CUM SE CONSTRUIEȘTE O MINTE

SECRETUL DEZVĂLUIT AL GÂNDIRII UMANE

RAY KURZWEIL

PARALELA 45
EDITURA

TOTAL

Editor: Călin Vlasie

Redactor: Ioan Es. Pop
Tehnoredactare: Gabriela Chircea
Coperta colecției: Ionuț Broștianu
Prepress: Marius Badea

Descrierea CIP a Bibliotecii Naționale a României
KURZWEIL, RAY
Cum se construieşte o minte: secretul dezvăluit al gândirii umane / Ray Kurzweil; trad.: Bogdan Chircea. - Piteşti: Paralela 45, 2013

ISBN 978-973-47-1651-7

I. Chircea, Bogdan (trad.)

004

Fotografia lui Rosalind Franklin de la pag. 22 este folosită cu acordul
National Portrait Gallery Londra:
NPG x76929
Rosalind Elsie Franklin
by Elliott & Fry
half-plate film negative, 11 June 1946
© National Portrait Gallery, London

Ray Kurzweil
How to create a mind. The secret of human thought revealed

Copyright © Ray Kurzweil, 2012

Published by arrangement with Viking, a member of Penguin Group (USA) Inc.
All rights reserved including the right of reproduction in whole or in part in any form.

Copyright© Editura Paralela 45, 2013, pentru prezenta ediție

Ray Kurzweil

CUM SE CONSTRUIEŞTE O MINTE

Secretul dezvăluit al gândirii umane

Traducere din limba engleză:
Bogdan Chircea

PARALELA 45

Îi dedic această carte tânărului Kurzweil al României, Ionuț Budișteanu, laureat al Premiului Turing – Nobelul informaticii. Pentru geniul lui și pentru speranța pe care mi-o dă în afirmarea și împlinirea posibilităților fără limite ale inteligenței omului. Și pentru că mi-a dat ocazia să văd pe viu ce înseamnă o minte cu adevărat strălucită.

Bogdan Chircea

Dedicație
Lui Leo Oscar Kurzweil.
Intri într-o lume extraordinară.

MULȚUMIRI

Vreau să-mi exprim gratitudinea față de soția mea, Sonya, pentru răbdarea ei plină de iubire în fața vicisitudinilor procesului creator;

Față de copiii mei, Ethan şi Amy, de nora mea, Rebecca, de sora mea, Enid, şi de nepotul meu, Leo, pentru iubire şi inspirație;

Față de mama mea, Hannah, pentru sprijinirea ideilor şi invențiilor mele de la începuturi, care mi-au dat libertatea să experimentez de la o vârstă fragedă şi pentru că l-a ținut pe tatăl meu în viață pe parcursul îndelungatei lui suferințe;

Față de editorul meu de multă vreme de la Viking, Rick Kot, pentru spiritul lui de conducător, pentru călăuzirea permanentă şi plină de pătrundere şi pentru redactarea expertă;

Față de Loretta Barrett, agentul meu literar de douăzeci de ani, pentru îndrumarea abilă şi plină de entuziasm;

Față de Aaron Kleiner, vechiul meu partener de afaceri, pentru colaborarea plină de devotament din ultimii patruzeci de ani;

Față de Amara Angelica, pentru sprijinul devotat şi excepțional pe care mi l-a acordat în munca de documentare;

Față de Sarah Black, pentru intuiția şi ideile extraordinare privitoare la documentare;

Față de Laksman Frank, pentru ilustrațiile de excepție;

Față de Sarah Reed, pentru sprijinul entuziast privitor la organizare;

Față de Nanda Barker-Hook, pentru organizarea la nivel de expert a evenimentelor mele publice pe această temă, dar şi pe altele;

Față de Amy Kurzweil, pentru îndrumarea în meşteşugul scrisului;

Față de Cindy Mason, pentru sprijinul în documentare, pentru ideile din domeniul inteligenței artificiale şi pentru conexiunea minte-corp;

Față de Dileep George, pentru ideile lui pătrunzătoare şi pentru discuțiile pline de inspirație purtate prin e-mail sau altfel;

Față de Martine Rothblatt, pentru devotamentul față de toate tehnologiile pe care le discut în această carte şi pentru colaborarea noastră în dezvoltarea tehnologiilor din aceste domenii;

Față de echipa KurzweilAI.net, care m-a sprijinit din punctul de vedere al documentării şi logistic în acest proiect, inclusiv față de Aaron Kleiner, Amara Angelica, Bob Beal, Casey Beal, Celia Black-Brooks, Cindy Mason, Denise Scutellaro, Joan Walsh, Giulio Prisco, Ken Linde, Laksman Frank, Maria Ellis, Nanda Barker-Hook, Sandi Dube, Sarah Black, Sarah Brangan şi Sarah Reed;

Față de echipa devotată de la Viking Penguin, pentru expertiza pe care mi-a pus-o la dispoziție, inclusiv față de Clare Ferraro (preşedinte), Carolyn Coleburn (director de publicitate), Yen Cheong şi Langan Kingsley (agenți de publicitate), Nancy Sheppard (director de marketing), Bruce Giffords (editor de producție), Kyle Davis (asistent editorial), Fabiana Van Arsdell (director de producție), Roland Ottewell (copy editor), Daniel Lagin (designer) şi Julia Thomas (designerul supracopertei);

Față de colegii mei de la Singularity University, pentru ideile, entuziasmul şi energia antreprenorială;

Față de colegii care mi-au dat idei generatoare de inspirație reflectate în această carte, inclusiv față de Barry Ptolemy, Ben Goertzel, David Dalrymple, Dileep George, Felicia Ptolemy, Francis Ganong, George Gilder, Larry Janowitch, Laura Deming, Lloyd Watts, Martine Rothblatt, Marvin Minsky, Mickey Singer, Peter Diamandis, Raj Reddy, Terry Grossman, Tomaso Poggio şi Vlad Sejnoha;

Față de lectorii experți, inclusiv față de Ben Goertzel, David Gamez, Dean Kamen, Dileep George, Douglas Katz, Harry George, Lloyd Watts, Martine Rothblatt, Marvin Minsky, Paul Linsay, Rafael Reif, Raj Reddy, Randal Koene, dr. Stephen Wolfram şi Tomaso Poggio;

Față de cititorii mei apropiați şi nespecialişti ale căror nume apar mai sus;

Şi, în cele din urmă, față de toți gânditorii creativi din lume care mă inspiră zi de zi.

INTRODUCERE

Creierul – mai mare decât cerul –
Căci – dacă le-aşezi alături –
Unul pe celălalt îl va cuprinde
Cu uşurinţă – cu tine – alături –
Creierul e mai adânc decât marea –
Căci – uite-le – albastru pe albastru –
Unul pe celălalt se înghit –
Ca bureţii – cofele –
Creierul are întocmai greutatea lui Dumnezeu –
Căci – cumpăneşte-le – gram cu gram –
De s-or deosebi – în caz c-o fi –
Ca de sunet, silaba.

— Emily Dickinson

Fiind cel mai important fenomen din Univers, inteligenţa poate transcende limitele naturale şi poate transforma lumea după propria imagine. În mâinile noastre, inteligenţa ne-a permis să ne depăşim limitările moştenirii biologice şi să ne schimbăm totodată pe noi înşine. Suntem singura specie care face acest lucru.

Povestea inteligenţei omului începe cu un univers capabil să codifice informaţia. Acesta a fost elementul care a facilitat evoluţia. Cum a ajuns Universul să arate astfel este în sine o poveste interesantă. Modelul-standard al fizicii are zeci de constante care trebuie să fie exact cele ce sunt, pentru că, altminteri, atomii n-ar fi fost posibili şi nu ar fi existat nici stele, nici planete, nici creier şi nici cărţi despre creier.

Faptul că legile fizicii sunt atât de precis reglate încât să permită evoluţia informaţiei pare să fie teribil de improbabil. Şi totuşi, potrivit principiului antropic, nu am vorbi despre ele dacă nu ar fi fost cazul.

Acolo unde unii oameni văd o intervenţie divină, alţii văd un multivers care susţine dezvoltarea de universuri, în timp ce universurile plictisitoare (nepurtătoare de informaţie) mor. Dar, indiferent de felul în care a ajuns

Universul aşa cum este, ne putem începe povestea cu o lume bazată pe informaţie.

Povestea evoluţiei se dezvăluie a fi plină de tot mai multe niveluri de abstracţie. Atomii – în special atomii de carbon, care pot să creeze structuri bogate în informaţie, legându-se pe patru direcţii diferite – au format molecule din ce în ce mai complexe. Ca urmare, fizica a dat naştere chimiei.

Un miliard de ani mai târziu, a apărut o moleculă complexă, numită ADN, care poate să codifice cu precizie şiruri lungi de informaţii şi să genereze organismele descrise de aceste „programe". Drept rezultat, chimia a dat naştere biologiei.

Cu o rată de dezvoltare tot mai rapidă, organismele şi-au format şi perfecţionat reţele de comunicare şi decizie numite sisteme nervoase, ca să poată coordona părţile din ce în ce mai complexe ale corpurilor lor şi comportamentele care le-au facilitat supravieţuirea. Neuronii care compun sistemele nervoase s-au agregat în creiere capabile de comportamente din ce în ce mai inteligente. În acest fel, biologia a dat naştere neurologiei, creierele fiind în această privinţă vârful de lance al stocării şi manipulării informaţiei. Şi astfel am ajuns de la atomi şi molecule la ADN şi creiere. Următorul pas a fost în mod unic omenesc.

Creierul mamiferelor are o aptitudine distinctă, nemaiîntâlnită la nicio altă clasă de animale. Suntem capabili de gândire *ierarhică*, de înţelegerea unei structuri compuse din diverse elemente aranjate într-un model, reprezentând acest aranjament cu un simbol şi apoi folosind acel simbol ca element al unei o configuraţii şi mai complicate. Această capacitate îşi are sediul într-o structură a creierului numită neocortex, care la om a atins un asemenea prag de complexitate, încât suntem capabili să denumim aceste modele *idei*. Printr-un proces recursiv fără sfârşit, suntem în stare să construim idei din ce în ce mai sofisticate. Numim această vastă multitudine de idei legate recursiv între ele *cunoaştere*. Numai *Homo sapiens* are o bază de cunoaştere care evoluează ea însăşi, se dezvoltă exponenţial şi e transmisă de la o generaţie la alta.

Dar creierele noastre au mai dat naştere unui nivel de abstracţie, concretizat în faptul că ne-am folosit inteligenţa, plus încă un factor de potenţare, un apendice opozabil – degetul mare –, pentru manipularea mediului în scopul confecţionării de unelte. Aceste unelte au reprezentat o nouă formă de evoluţie, neurologia dând naştere tehnologiei. Numai datorită uneltelor, baza noastră de cunoaştere a putut să se dezvolte fără limită.

Prima noastră invenție a fost povestea: limbajul vorbit, care ne-a permis să reprezentăm idei prin enunțuri distincte. Odată cu inventarea ulterioară a limbajului scris, ne-am dezvoltat forme distincte pentru a ne simboliza ideile. Bibliotecile cu limbaj scris au depășit cu mult capacitatea creierului nostru neasistat de a reține și a extinde baza de cunoaștere de idei structurate recursiv.

Există o anumită controversă dacă alte specii, cum fi cimpanzeii, au capacitatea de exprimare a ideilor ierarhice prin limbaj. Cimpanzeii sunt capabili să învețe un set limitat de simboluri ale limbajului semnelor, pe care le pot folosi pentru comunicarea cu antrenorii umani. Este limpede, pe de altă parte, că există limite clare în ceea ce privește complexitatea structurilor de cunoaștere cu care sunt capabili să opereze cimpanzeii. Propozițiile pe care le pot exprima sunt limitate la anumite secvențe specifice simple substantiv-verb și nu sunt capabili de expansiunea nelimitată a complexității, caracteristică omului. Ca exemplu distractiv de complexitate a limbajului generat de om, citește una dintre propozițiile care se întind pe mai multe pagini dintr-o povestire sau nuvelă de Gabriel García Márquez. Povestea întinsă pe șase pagini *Ultima călătorie a fantomei* constă dintr-o singură frază care curge foarte bine atât în limba spaniolă, cât și în traducerea engleză.[1]

Ideea fundamentală a celor trei cărți anterioare ale mele pe tema tehnologiei (*Epoca mașinilor inteligente*, scrisă în anii 1980 și publicată în 1989; *Epoca mașinilor spirituale*, scrisă la mijlocul anilor 1990 și publicată în 1999; și *Singularitatea e aproape*, scrisă la începutul anilor 2000 și publicată în 2005) este că procesul evolutiv se accelerează în mod inerent (ca urmare a nivelului tot mai ridicat de abstractizare) și că produsele lui se dezvoltă exponențial, atât sub aspectul complexității, cât și sub cel al capacității. Eu numesc acest fenomen *Legea accelerării veniturilor,* valabilă și în evoluția biologică, și în cea tehnologică. Exemplul cel mai spectaculos de aplicare a Legii accelerării veniturilor îl reprezintă creșterea exponențială remarcabil de predictibilă a capacității și raportului preț/performanță în tehnologia informației. Procesul evolutiv al tehnologiei a condus invariabil la computer, care la rândul lui a făcut posibilă o vastă extindere a bazei noastre de cunoaștere, permițând realizarea unor conexiuni vaste între un domeniu al cunoașterii și altul. Web-ul este intrinsec o expresie elocventă și adecvată a capacității unui sistem ierarhic de a integra o mare cantitate de cunoaștere, în condițiile păstrării structurii sale inerente. Lumea însăși este inerent ierarhică. De

pildă, copacii au ramuri; ramurile au frunze; frunzele au vinişoare. Clădirile au etaje; etajele au încăperi; încăperile au uşi, ferestre, ziduri şi pardoseli.

De asemenea, am pus la punct unelte care ne permit acum să ne înţelegem propria biologie în termeni de informaţie precişi. Prin retroinginerie, prelucrăm rapid procesele informaţiei aflate la baza biologiei, inclusiv pe cele ale creierului nostru. Avem actualmente codul-obiect al vieţii sub forma genomului uman, o realizare care a constituit prin ea însăşi un remarcabil exemplu de creştere exponenţială, prin faptul că volumul de date genetice secvenţializate aproape că s-a dublat în lume în fiecare an pe parcursul ultimelor două decenii.[2] Astăzi, avem capacitatea de a simula pe computere modul cum dau naştere secvenţele de perechi de baze secvenţelor de aminoacizi ce se adună în proteinele tridimensionale din care este construită întreaga lume biologică. Complexitatea proteinelor pentru care putem simula concentrarea a crescut continuu, pe măsură ce resursele computaţionale au manifestat o creştere exponenţială.[3] Putem de asemenea simula modul în care interacţionează proteinele între ele într-un dans tridimensional de forţe atomice extrem de complicat. Înţelegerea tot mai aprofundată a biologiei de către om este una dintre cele mai importante faţete ale descoperirii secretelor inteligente pe care ni le-a transmis evoluţia. Apoi, folosind aceste paradigme inspirate de biologie, urmează crearea unei tehnologii şi mai inteligente.

Actualmente este în curs de derulare un proiect grandios, care antrenează mii şi mii de oameni de ştiinţă şi ingineri. Aceştia se străduiesc să înţeleagă cel mai inteligent exemplu de proces pe care îl avem la dispoziţie: creierul omului. Este în mod justificat cel mai important efort din istoria civilizaţiei om-maşină. În cartea mea *The Singularity Is Near (Singularitatea este aproape)*, am demonstrat că un corolar al Legii accelerării veniturilor afirmă că s-ar putea să nu existe alte specii inteligente. Ca să rezum demonstraţia: dacă aceste specii ar fi existat, le-am fi observat, dat fiind intervalul de timp relativ redus care se scurge între momentul achiziţiei de către o civilizaţie a tehnologiei brute (să ne gândim că, în 1850, cel mai rapid mijloc de transmitere a informaţiei la nivel naţional, în Statele Unite, era poşta călare – Pony Express) şi momentul deţinerii acelei tehnologii care poate depăşi graniţele propriei planete.[4] Din această perspectivă, supunerea creierului uman la mijloacele retroingineriei poate fi considerată cel mai proiect din Univers.

Scopul proiectului este înţelegerea precisă a modului cum funcţionează creierul uman, urmată de utilizarea metodelor scoase la lumină pentru a ne înţelege mai bine pe noi înşine, pentru repararea creierului la nevoie şi – cel

mai relevant în privința subiectului acestei cărți – pentru crearea unor mașini și mai inteligente. Nu uita, cititorule, că ingineria este capabilă tocmai de amplificarea unui fenomen natural. Ca exemplu, să ne gândim la Principiul lui Bernoulli, care afirmă că există o presiune ceva mai redusă pe o suprafață curbă în mișcare, decât pe una dreaptă. Aspectele matematice ale modului în care Principiul lui Bernoulli descrie portanța aripii nu sunt încă definitiv stabilite în rândurile oamenilor de știință, dar ingineria a preluat această idee delicată, și-a concentrat forțele și a creat întreaga lume a aviației.

În această carte, prezint o teză pe care o numesc Teoria minții bazate pe recunoașterea modelelor – *Pattern recognition theory of mind (PRTM)* – care, după cum demonstrez, descrie algoritmul de bază al neocortexului (regiunea creierului răspunzătoare de percepție, memorie și gândire critică). În capitolele care urmează, prezint modul în care cercetările recente din domeniul neuroștiinței, dar și experimentele noastre mintale, conduc la concluzia inevitabilă că această metodă e folosită total la nivelul neocortexului. Implicațiile Teoriei minții bazate pe recunoașterea modelelor, combinată cu Legea accelerării veniturilor, spun că vom fi capabili să valorificăm prin inginerie aceste principii pentru extinderea extraordinară a puterii propriei noastre inteligențe.

Iar acest proces este deja în curs de desfășurare. Există o multitudine de funcții și activități – care anterior reprezentau apanajul exclusiv al inteligenței umane – care acum pot fi realizate de computere, de obicei cu mai multă precizie și pe o scară mult mai mare. De fiecare dată când trimiți un e-mail sau dai un apel pe un telefon celular, algoritme inteligente rulează în mod optim informația. Faci o electrocardiogramă, iar aceasta îți este furnizată împreună cu un diagnostic computerizat care rivalizează cu diagnosticele medicilor. Același lucru este valabil și în ceea ce privește imaginile elementelor figurate ale sângelui. Algoritmi inteligenți detectează automat fraudele cu carduri bancare, pilotează și aduc la sol avioane, ghidează sisteme de arme inteligente, ajută la conceperea unor produse prin proiectare asistată de computer și urmăresc nivelurile de inventar la momentul oportun, asamblează produse în fabrici robotizate și joacă diverse jocuri, cum ar fi șahul și chiar subtilul Go, la nivel de maeștri.

Milioane de oameni au fost martorii momentului când un computer IBM pe nume Watson a concurat la *Jeopardy!*, un joc bazat pe limbajul natural, și a obținut un scor mai mare decât cei mai buni doi jucători umani din lume la un loc. Ar trebui subliniat nu numai faptul că Watson a citit și a „înțeles"

limbajul subtil al întrebărilor de la *Jeopardy!* (care include fenomene cum ar fi jocurile de cuvinte şi metaforele), ci şi faptul că şi-achiziţionat singur cunoştinţele necesare pentru a da răspunsurile, asimilând sute de milioane de pagini de documente în limbaj natural, inclusiv Wikipedia şi alte enciclopedii. A trebuit să stăpânească practic toate domeniile care ilustrează strădaniile intelectuale ale omului, inclusiv istoria, ştiinţele, literatura, artele, cultura şi multe altele. Compania IBM lucrează actualmente, împreună cu Nuance Speech Technologies (anterior Kurzweil Computer Products, prima mea companie), la o nouă versiune a lui Watson, care va citi literatură medicală (în esenţă, toate jurnalele şi blogurile medicale importante) – pentru a deveni maestru al diagnosticului şi consultant medical – folosind limbajul clinic Nuance: înţelegerea tehnologiilor. Unii observatori au argumentat că, de fapt, Watson nu „înţelege" cu adevărat întrebările de la *Jeopardy!* sau enciclopediile pe care le-a citit şi că realizează doar o „analiză statistică". Un aspect-cheie pe care îl voi descrie aici priveşte faptul că tehnicile matematice dezvoltate în domeniul inteligenţei artificiale (cum ar fi cele folosite de Watson şi Siri, asistentul iPhone) sunt foarte asemănătoare din punct de vedere matematic cu metodele pe care le-a dezvoltat biologia sub forma neocortexului. Dacă înţelegerea limbajului şi a altor fenomene prin analiză statistică nu explică adevărata înţelegere, atunci nici oamenii nu sunt înzestraţi cu putere de înţelegere.

Capacitatea lui Watson de a stăpâni în mod inteligent cunoaşterea din documente în limbaj natural se va regăsi pe un motor de căutare pe care îl vei avea şi tu la dispoziţie – şi asta chiar în curând. Deja oamenii vorbesc cu telefoanele în limbaj natural (prin intermediul lui Siri, la care a contribuit şi Nuance). Aceşti asistenţi în limbaj natural vor deveni rapid mai inteligenţi, pe măsură ce vor folosi mai multe metode gen Watson. Watson însuşi va continua să se perfecţioneze.

Automobilele Google fără şofer au parcurs deja 320.000 de kilometri în oraşe mari şi mici din California (fără îndoială, cifra va fi cu mult mai mare când această carte va poposi pe rafturile reale şi virtuale). Există multe alte exemple de inteligenţă artificială în lumea de astăzi şi cu mult mai multe se întrezăresc la orizont.

Altă ilustrare a Legii accelerării veniturilor se referă la rezoluţia spaţială a scanării creierului. Volumul de date pe care le culegem despre creier se dublează de la an la an. În plus, demonstrăm că putem transforma aceste date în modele de lucru şi în simulări ale regiunilor creierului. Am reuşit să

reproducem prin retroinginerie funcții-cheie ale cortexului auditiv, unde procesăm informațiile despre sunete; ale cortexului vizual, unde procesăm informații de la analizatorul vizual; și ale cerebelului, cu care contribuim parțial la formarea abilităților (cum ar fi prinderea unei mingi).

În avangarda proiectului de înțelegere, modelare și simulare a creierului uman stă intenția de a reproduce prin retroinginerie neocortexul cerebral, în care realizăm gândirea ierarhică recursivă. Cortexul cerebral reprezintă 80 la sută din creierul omului și este format din structuri extrem de repetitive, care le permit oamenilor să creeze ansambluri de idei cu complexități arbitrare.

În cadrul Teoriei minții pe baza recunoașterii modelelor, descriu un model de achiziționare de către creierul uman a acestei capacități fundamentale folosind o structură foarte inteligentă, proiectată de evoluția biologică. În acest mecanism cortical există detalii pe care încă nu le înțelegem pe deplin, dar cunoaștem suficient de bine funcțiile pe care trebuie să le realizeze acesta pentru a putea să creăm algoritmi care să corespundă exact aceluiași scop. Începând să înțelegem neocortexul, ne aflăm acum în poziția de a-i amplifica foarte mult puterea, la fel cum lumea aviației a amplificat foarte mult puterea Principiului lui Bernoulli. Principiul de operare al neocortexului este în mod demonstrabil cea mai importantă idee din lume, deoarece e capabil să reprezinte integralitatea cunoașterii și aptitudinilor și să creeze noi cunoștințe. La urma urmelor, neocortexul a fost răspunzător de toate romanele, cântecele, picturile, descoperirile științifice, ca și de alte, felurite, produse ale gândirii omenești.

În domeniul neuroștiinței, este foarte necesară o lege care să facă legătura între observațiile extrem de disparate și extinse comunicate zilnic. O teorie unificată reprezintă cerința crucială în toate domeniile științei. În Capitolul 1, voi arăta cum doi visători au unificat biologia și fizica, domenii care anterior păruseră disperant de dezordonate și străine unul de altul, iar pe urmă mă voi ocupa de modul în care o asemenea teorie poate fi aplicată peisajului creierului.

Astăzi întâlnim adeseori grandioase manifestări ale complexității creierului uman. Google dă în jur de 30 de milioane de link-uri la o căutare care solicită citate pe această temă. (Este imposibil, însă, să se traducă asta în numărul real de citate pe care le afișează, deoarece unele site-uri web conțin citate multiple, pe când altele nu conțin nici unul.) James D. Watson însuși a scris în 1992: „Creierul este ultima și cea mai măreață frontieră a biologiei, cel mai complex lucru care a fost descoperit până acum în Universul nostru".

El continuă şi explică de ce crede că „[creierul] conţine sute de miliarde de celule interconectate prin intermediul a mii de miliarde de conexiuni. Creierul descumpăneşte mintea".[5] Sunt de acord cu sentimentul lui Watson referitor la faptul că creierul este cea mai măreaţă frontieră a biologiei, însă faptul că acesta conţine miliarde de celule şi mii de miliarde de conexiuni nu înseamnă că metoda lui principală este foarte complexă, dacă putem identifica modele uşor de înţeles (şi recreabile) în aceste celule şi conexiuni, mai ales în cele masiv redundante.

Să ne gândim la ce înseamnă să fii complex. Am putea să ne întrebăm: O *pădure e complexă?* Răspunsul depinde de perspectiva pe care alegem să o urmăm. Putem observa că în pădure sunt mii de copaci şi că fiecare dintre ei e diferit. Putem să mergem, pe urmă, mai departe şi să observăm că fiecare copac are mii de ramuri şi că fiecare ramură e total diferită de celelalte. Apoi, am putea trece la descrierea ciudăţeniilor unei singure ramuri. Concluzia poate fi aceea că pădurea prezintă o complexitate care depăşeşte şi cea mai prolifică imaginaţie.

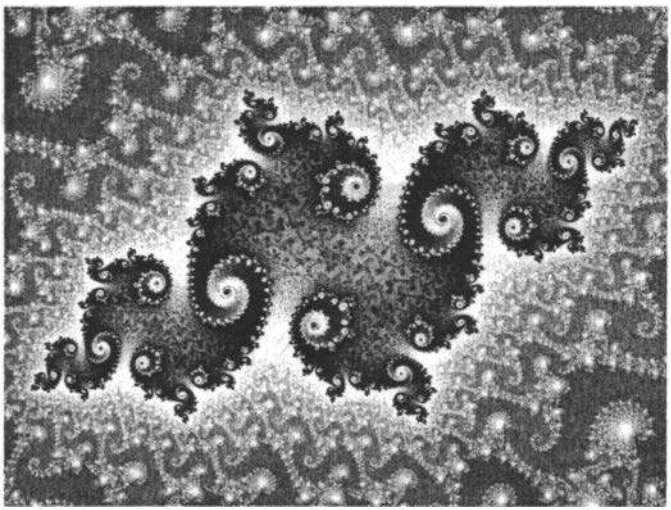
Însă o asemenea abordare ar însemna literalmente să nu vedem pădurea din cauza copacilor. Desigur, există o imensă variaţie fractală între copaci şi între ramuri, dar, pentru a înţelege corect principiile unei păduri, ar fi mai bine dacă am începe cu identificarea modelelor distincte de redundanţă cu variaţie stohastică (respectiv aleatorie) ce pot fi găsite în ea. Ar fi corect să spunem că acest concept de pădure este mai simplu decât conceptul de copac.

La fel stau lucrurile şi în cazul creierului, care prezintă, de asemenea, redundanţe enorme, mai ales la nivelul neocortexului. Aşa cum voi arăta în această carte, corect ar fi să existe o complexitate mai mare într-un singur neuron, decât în structura de ansamblu a neocortexului.

În mod cert, scopul meu în lucrarea de faţă nu este acela de a adăuga un nou citat la milioanele care există deja şi care atestă cât de complex este creierul; dimpotrivă, doresc să îţi imprim, cititorule, forţa simplităţii sale. Şi voi face acest lucru descriind modul în care un mecanism de bază ingenios, de recunoaştere, reţinere şi predicţie, repetat în neocortex de sute de milioane de ori, explică marea diversitate a gândirii noastre. La fel cum, din diferitele combinaţii ale valorilor codului genetic din ADN-ul nuclear şi mitocondrial, apare o uluitoare diversitate de organisme, tot aşa se formează şi o gamă uluitoare de idei, gânduri şi abilităţi pe baza valorilor modelelor (de conexiuni şi tării sinaptice) care se găsesc în şi între identificatorii noştri neocorticali

de modele. Așa cum spune specialistul în neuroştiință de la MIT Sebastian Seung, „identitatea nu stă în genele noastre, ci în conexiunile dintre celulele creierului".[6]

Trebuie să facem distincția între adevărata complexitate de concepție şi complexitatea aparentă. Să luăm celebrul set Mendelbrot, a cărui imagine este de multă vreme simbolul complexității. Pentru înțelegerea caracterului său aparent complicat, este suficient să mărim imaginea (pe care o poți accesa, cititorule, prin intermediul link-urilor din această notă de final).[7] Există o nesfârşită complexitate în complexitate şi toate sunt întotdeauna diferite. Cu toate acestea, concepția – formula – pentru setul Mendelbrot nu ar putea fi mai simplă de-atât. Are o lungime de şase caractere: $Z = Z^2 + C$, unde Z e un număr „complex" (ceea ce înseamnă o pereche de numere), iar C este o constantă. Nu-i necesar să înțelegi funcția Mendelbrot pentru a pricepe că e simplă.



Vedere a unei reprezentări a setului Mendelbrot, o formulă simplă care este aplicată recurent. Pe măsura măririi reprezentării pe ecran, imaginile se schimbă în permanență în moduri aparent complexe

Această formulă este aplicată recurent la fiecare nivel al unei ierarhii. Lucrul e valabil şi pentru creier. Structura lui repetitivă nu este la fel de simplă precum cea a formulei compuse din şase caractere a setului Mandelbrot, dar

nu e nici pe departe atât de complexă pe cât ar sugera-o milioanele de citate referitoare la complexitatea creierului. Acest model neocortical se repetă la nesfârşit la toate nivelurile ierarhiei conceptuale reprezentate de neocortex. Einstein a articulat bine scopurile cărţii de faţă când a spus că „orice nebun inteligent poate să facă lucrurile mai mari şi mai complexe... dar e nevoie... de mult curaj pentru a înainta în direcţia opusă".

Până acum am vorbit despre creier. Dar cum rămâne cu mintea? De exemplu, cum poate să ajungă la conştiinţă un neocortex care rezolvă probleme? Şi pentru că tot suntem la acest subiect, câte minţi conştiente avem în creierul nostru? Există mărturii care sugerează că e posibil să fie mai mult de una.

O altă întrebare pertinentă referitoare la minte sună aşa: *Ce este liberul arbitru – şi îl avem noi oare?* Există experimente care par să indice faptul că începem să ne implementăm deciziile încă înainte de a ne da seama că le-am luat deja. Asta să presupună faptul că liberul arbitru e o iluzie?

În sfârşit, care atribute ale creierului nostru sunt răspunzătoare pentru constituirea identităţii proprii? Sunt aceeaşi persoană care eram în urmă cu şase luni? În mod evident, nu mai sunt exact cum eram atunci, dar am aceeaşi identitate?

Vom trece în revistă ce presupune Teoria pentru minte a recunoaşterii modelelor în privinţa acestor întrebări vechi de secole.

Capitolul 1

EXPERIMENTE MINTALE DESPRE LUME

Teoria selecției naturale a lui Darwin a apărut foarte târziu în istoria gândirii. O fi fost oare întârziată pentru că se opunea dezvăluirii adevărului, pentru că era o temă cu totul nouă în istoria științei, pentru că era caracteristică numai ființelor vii sau pentru că se ocupa cu scopul și cauzele finale fără să postuleze un act al creației? Eu cred că nu. Darwin pur și simplu a descoperit rolul selecției, un gen de cauzalitate foarte diferit de mecanismele cauză-efect ale științei de până în acel moment.
Originile unei fantastice varietăți de lucruri vii pot fi explicate prin contribuția acelor caracteristici noi, posibil de proveniență aleatorie, care au reușit să supraviețuiască. În științele fizicii și biologiei exista prea puțin sau nu exista nimic care să prevestească selecția drept cauză principală.

— B.F. Skinner

În ultimă instanță, nimic nu este sacru, cu excepția integrității propriei minți.

— Ralph Waldo Emerson

O metaforă din geologie

La începutul secolului al nouăsprezecelea, geologii au meditat la o întrebare fundamentală. Peșterile și canioanele de felul Marelui Canion din Statele Unite și al Defileului Vikos din Grecia (recunoscut ca fiind cel mai adânc canion din lume) există în toate colțurile lumii. Cum au ajuns acolo aceste formațiuni maiestuoase?

Invariabil, a existat un curs de apă care a părut să profite de oportunitatea de a curge prin aceste structuri naturale, dar, anterior mijlocului de secol nouăsprezece, păruse absurd ca aceste cursuri de apă blânde să fi fost creatoarele unor văi și promontorii atât de mari. Însă geologul britanic Charles Lyell (1797-1875) a avansat ipoteza că într-adevăr deplasarea apei a fost aceea care a sculptat aceste modificări geologice majore, de-a lungul

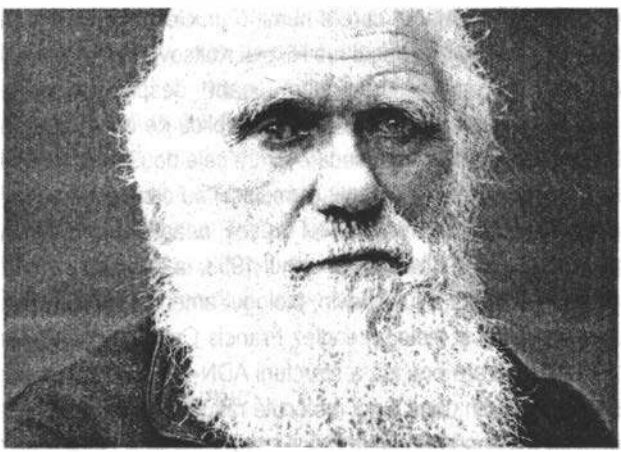
unor perioade îndelungate, în esență grăunte cu grăunte. Ideea lui Lyell a fost inițial întâmpinată cu ridiculizări, dar, în decurs de două decenii, teza a căpătat susținere generală.

Un om care urmărea cu atenție răspunsul comunității științifice la teza radicală a lui Lyell a fost naturalistul englez Charles Darwin (1809-1882). Să ne gândim la situația biologiei în jurul anului 1850. Domeniul era extraordinar de stufos, se confrunta cu nenumărate specii de animale și plante, oricare dintre ele prezentând o mare complexitate. Majoritatea oamenilor de știință s-au opus oricărei încercări de a veni cu o teorie unificatoare a uluitoarei varietăți a Naturii. Această diversitate slujea drept testament pentru glorificarea creației lui Dumnezeu, ca să nu mai vorbim de inteligența oamenilor de știință capabili să o stăpânească.

Darwin a abordat problema conceperii unei teorii generale a speciilor făcând analogie cu teza lui Lyell pentru explicarea modificărilor graduale ale caracteristicilor speciilor pe parcursul mai multor generații. El a combinat această idee cu propriile sale experimente mintale și cu observațiile de pe parcursul celebrei lui lucrări *Călătoria corabiei Beagle*. Darwin a demonstrat că, de-a lungul fiecărei generații, indivizii care pot supraviețui cel mai bine în nișa lor ecologică ar fi și cei care ar crea următoarea generație.

În data de 22 noiembrie 1859, cartea lui Darwin *Originea speciilor* a fost pusă în vânzare. În ea, autorul își exprima clar datoria față de Lyell:

> Sunt perfect conștient că această doctrină a selecției naturale, exemplificată în anterioarele postùri imaginare, este expusă aceloraşi obiecții care au fost emise la început și împotriva nobilelor perspective ale lui Sir Charles Lyell cu privire la „modificările moderne ale Pământului, ilustrate de geologie"; dar acum rareori mai auzim că acțiunea valurilor de pe coaste mai este considerată o cauză superficială și nesemnificativă atunci când e aplicată excavării unor văi gigantice sau formării celor mai lungi linii de promontorii din interiorul uscatului. Selecția naturală nu poate să intervină decât prin acumularea și păstrarea unor modificări infinit de mărunte, fiecare dintre ele profitabilă pentru ființa actuală; și, la fel cum geologia modernă aproape că a alungat concepții cum ar fi excavarea unei mari văi de un singur val de proporții diluviene, tot așa selecția naturală, dacă este un principiu adevărat, va elimina credința în crearea permanentă de noi ființe organice sau în orice mari și bruște modificări ale structurii lor.[1]



Charles Darwin, autorul cărții *Originea speciilor*, care a consacrat ideea de evoluție biologică

Întotdeauna există mai multe motive pentru care se manifestă opoziție față de marile idei noi şi nu este greu să identificăm asemenea motive şi în cazul lui Darwin. Afirmația că ne-am tras nu de la Dumnezeu, ci din maimuțe, iar înaintea acestora din viermi, nu le-a căzut bine multor comentatori. Aluzia că, bunăoară, câinele nostru de companie ne este văr, la fel ca şi coropişnița, ca să nu mai vorbim de planta pe care am călcat (văr de gradul un milion sau un miliard, poate, dar totuşi rudă) li s-a părut multora o blasfemie.

Însă ideea a prins rapid, pentru că aducea coerență în ceea ce anterior reprezentase un corp de observații aparent fără legătură între ele. Înainte de anul 1872, la publicarea celei de-a şasea ediții a cărții *Originea speciilor*, Darwin a adăugat într-un pasaj: „Ca o consemnare a unei stări anterioare de lucruri, am păstrat în paragrafele precedente... mai multe fraze care sugerează că naturaliştii cred într-o creație separată a fiecărei specii; şi am fost foarte cenzurat pentru că m-am exprimat astfel. Dar, fără îndoială, aceasta era credința generală atunci când a apărut prima ediție a prezentei lucrări. Acum lucrurile s-au schimbat complet şi aproape orice naturalist recunoaşte marele principiu al evoluției".[2]

Pe parcursul următorului secol, ideea unificatoare a lui Darwin s-a adâncit. În anul 1869, la doar un deceniu de la publicarea inițială a *Originii speciilor*, fizicianul elvețian Friedrich Miescher (1844-1895) a descoperit în

nucleul celulei o substanță pe care a numit-o „nucleină" şi care s-a dovedit a fi ADN-ul.[3] În anul 1927, biologul rus Nikolai Koltsov (1872-1940) a descris ceea ce el numea o „moleculă ereditară gigant", despre care a spus că e formată din „două şiruri în oglindă, care se replică de o manieră semiconservatoare, folosind drept model fiecare dintre cele două şiruri". Şi descoperirea lui a fost condamnată de mulți. Comuniştii au considerat că era vorba de propagandă nazistă, iar moartea lui bruscă, neaşteptată, a fost atribuită poliției secrete a Uniunii Sovietice.[4] În anul 1953, la aproape un secol de la publicarea prolificei cărți a lui Darwin, biologul american James D. Watson (născut în anul 1928) şi biologul englez Francis Crick (1916-2004) au realizat prima caracterizare precisă a structurii ADN-ului, descriindu-l drept un dublu helix compus din două lungi molecule răsucite.[5] Merită să scot în evidență faptul că descoperirea celor doi s-a bazat pe ceea ce actualmente e cunoscut sub numele de „Fotografia 51", realizată de colega lor, Rosalind Franklin, care a folosit cristalografia în raze X. A fost prima reprezentare care a dezvăluit dublul helix. Date fiind indiciile obținute din fotografia lui Franklin, au fost emise sugestii potrivit cărora Rosalind Franklin ar fi trebuit să primească, alături de Watson şi Crick, Premiul Nobel.

Rosalind Franklin a realizat fotografia esențială a ADN-ului (folosind cristalografia în raze X), care le-a permis lui Watson şi Crick să descrie cu acuratețe, pentru prima oară, structura ADN-ului

Odată cu descrierea moleculei care putea să codifice programul biologic, a fost instaurată şi o teorie unificatoare a biologiei. Aceasta a asigurat un fundament simplu şi elegant pentru toate formele de viață. Depinzând numai de valorile perechilor de bază care alcătuiesc şirurile ADN-ului din nucleu (şi, într-o mai mică măsură, din mitocondrii), organismul devine la maturitate un fir de iarbă sau o ființă omenească. Intuirea acestui fapt nu a eliminat încântătoarea diversitate a Naturii, dar acum înțelegem că această extraordinară diversitate izvorăşte din marea varietate de structuri care pot fi codificate în această moleculă universală.

Călare pe o rază de lumină

La începutul secolului douăzeci, lumea fizicii se afla într-o fundătură, în urma unei noi serii de experimente. În anul 1879, în familia unui inginer german şi a soției sale casnice s-a născut un băiat. A început să vorbească abia la vârsta de trei ani şi s-a spus despre el că, la nouă ani, a avut probleme la şcoală. La şaisprezece ani, visa cu ochii deschişi să călărească pe o rază de lumină.

Acest băiat era conştient de experimentul din anul 1803 al matematicianului englez Thomas Young (1773-1829), care a stabilit că lumina e formată din unde. Concluzia, la momentul acela, a fost că undele de lumină trebuie să se deplaseze printr-un fel de mediu; la urma urmelor, valurile oceanului se propagă prin apă, iar undele sonore se propagă prin aer şi alte materiale. Oamenii de ştiință au denumit mediul prin care se propagă lumina *eter*. Băiatul era la curent şi cu experimentul din anul 1887 al savanților americani Albert Michelson (1852-1931) şi Edward Morley (1838-1923), care încercaseră să confirme existența eterului. Acel experiment se bazase pe analogia deplasării unei bărci cu vâsle în amonte şi în aval pe un râu. Dacă vâsleşti cu viteză constantă, atunci viteza ta măsurată de pe mal va fi mai mare când vâsleşti în aval decât când vâsleşti în amonte. Michelson şi Morley au presupus că lumina se deplasează prin eter cu o viteză constantă (respectiv cu viteza luminii). Ei au raționat că viteza luminii Soarelui atunci când Pământul se deplasează pe orbita lui spre Soare (viteză măsurată din punctul nostru de vedere, de pe Pământ) comparativ cu viteza ei aparentă când Pământul se depărtează de Soare trebuie să fie diferită (cu dublul vitezei Pământului). Dovedirea acestui lucru ar fi confirmat existența eterului. Însă ceea ce au descoperit cei

doi americani a fost faptul că nu există nicio diferență între vitezele cu care lumina solară ajunge pe Pământ, indiferent unde se află acesta pe orbita lui. Descoperirea infirma ideea de eter, dar ce se întâmpla în realitate? Acest lucru a rămas un mister timp de aproape două decenii.

În timp ce își imagina că se plimbă de-a lungul unei raze, adolescentul german s-a gândit că trebuia să vadă razele de lumină nemișcate, la fel cum un tren ar părea că nu se mișcă deloc dacă te-ai deplasa alături de el cu aceeași viteză. Însă și-a dat seama că acest lucru era imposibil, deoarece se presupunea că viteza luminii e constantă indiferent de mișcarea observatorului. Prin urmare, și-a imaginat că se deplasează de-a lungul unei raze de lumină, dar la o viteză întrucâtva mai mică. Și dacă se deplasa cu 90 la sută din viteza luminii? Dacă razele de lumină sunt asemenea trenurilor, a raționat el, atunci trebuie să vadă raza de lumină deplasându-se cu 10 la sută din viteza luminii. Și, într-adevăr, asta ar fi trebuit să vadă observatorii de pe Pământ. Însă știm că viteza luminii este o constantă, așa cum arătase experimentul Michelson-Morley. Prin urmare, tânărul ar fi văzut în mod necesar raza de lumină deplasându-se înaintea lui cu întreaga viteză a luminii. Acest fapt părea o contradicție – deci cum putea fi posibil?

Răspunsul i-a devenit evident germanului, al cărui nume era, întâmplător, Albert Einstein (1879-1955), înainte să împlinească douăzeci și șase de ani. În mod evident – pentru tânărul maestru Einstein – *însuși timpul ar fi trebuit să încetinească pentru el*. Einstein își explică raționamentul într-o lucrare publicată în anul 1905.[7] Dacă observatorii de pe Pământ ar fi putut să se uite la ceasul tânărului, l-ar fi văzut ticăind de zece ori mai încet. Și pe bună dreptate, la întoarcerea pe Pământ, ceasul lui ar fi arătat că s-au scurs numai 10 procente din timp (ignorând, pentru moment, accelerarea și decelerarea). Din perspectiva lui, însă, ceasul ar funcționa normal și raza de lumină de lângă el s-ar deplasa cu viteza luminii. Încetinirea de zece ori a vitezei timpului însuși (relativ la ceasurile de pe Pământ) explică pe deplin aparentele discrepanțe de perspectivă. La extrem, încetinirea trecerii timpului ar ajunge la zero în momentul în care viteza deplasării ar atinge viteza luminii; prin urmare, era imposibil să se deplaseze odată cu raza de lumină. Cu toate că deplasarea cu viteza luminii era imposibilă, s-a dovedit a nu fi teoretic imposibilă deplasarea cu viteze *superioare* celei a razei de lumină. În acest caz, timpul ar curge invers.

Acest deznodământ li s-a părut absurd multora dintre criticii inițiali. Cum putea timpul să își încetinească curgerea doar în funcție de viteza deplasării

cuiva? Şi, într-adevăr, timp de optsprezece ani de la momentul experimentului Michelson-Morley, alți gânditori nu reuşiseră să găsească o concluzie care i se părea atât de evidentă Maestrului Einstein. Acei mulți alții care cumpăniseră această problemă în ultima parte a secolului al nouăsprezecelea practic „căzuseră de pe cal" în privința urmăririi implicațiilor unui principiu, ținându-se în schimb de ideile lor preconcepute privitoare la mecanismul realității. (Probabil că ar fi trebuit să modific metafora în „căzuseră de pe raza de lumină".)

Al doilea experiment mintal al lui Einstein a fost să se imagineze pe el şi pe fratele său zburând prin spațiu. Cei doi se află la 297.000 de kilometri distanță unul de altul. Einstein vrea să se deplaseze mai repede, dar doreşte şi să păstreze distanța dintre ei constantă. Aşadar îi face fratelui său un semn cu lanterna de fiecare dată când vrea să accelereze. Deoarece ştie că va fi nevoie de o secundă pentru ca semnalul să ajungă la fratele lui, aşteaptă timp de o secundă după emiterea semnalului înainte să efectueze propria accelerare. De fiecare dată când recepționează semnalul, fratele lui accelerează imediat. În acest fel, cei doi frați accelerează exact în acelaşi timp şi, prin urmare, rămân la aceeaşi distanță unul de celălalt.

Dar acum să ne gândim la ce am vedea dacă ne-am afla pe Pământ. Dacă frații s-ar îndepărta de noi (Albert fiind în față), ni s-ar părea că ar dura mai puțin de o secundă pentru ca lumina să ajungă la fratele lui, pentru că el se îndreaptă către sursa de lumină. De asemenea, am vedea ceasul fratelui lui Albert încetinind (pe măsură ce viteza lui creşte şi el se află mai aproape de noi). Din ambele motive, i-am vedea pe amândoi frații apropiindu-se din ce în ce mai mult şi în cele din urmă ciocnindu-se. Însă, din perspectiva lor proprie, cei doi frați ar rămâne la o distanță constantă de 297.000 de kilometri unul de celălalt.

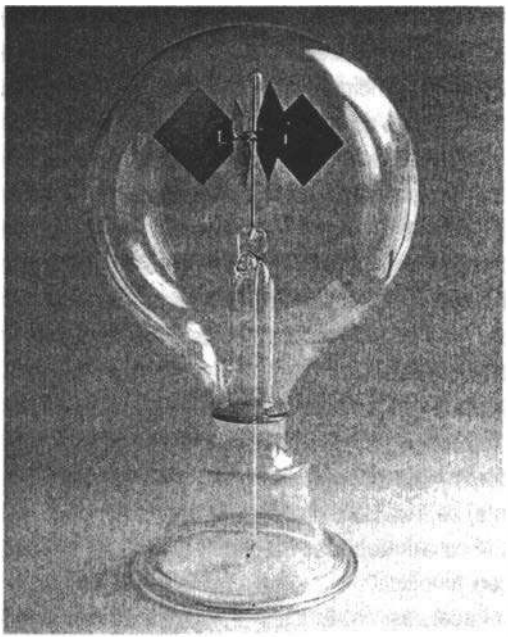
Cum e posibil aşa ceva? Răspunsul – *evident* – este că distanțele se contractă în paralel cu deplasarea (dar nu şi perpendicular pe ea). Prin urmare, cei doi frați Einstein ar deveni mai scurți (presupunând că ar zbura cu capul înainte) pe măsură ce s-ar deplasa tot mai rapid. Această concluzie bizară probabil că a îndepărtat mai mulți fani inițiali ai lui Einstein decât diferența trecerii timpului.

În cursul aceluiaşi an, Einstein s-a gândit la relația dintre materie şi energie. Era un alt experiment mintal. Fizicianul scoțian James Clerk Maxwell arătase în anii 1850 că particulele de lumină, numite fotoni, nu aveau masă, dar cu toate acestea aveau impuls. Când eram mic, am avut un dispozitiv

numit radiometru Crookes[8], care consta dintr-o sferă de sticlă închisă ermetic, în care era un vid parțial şi un set de patru pale ce se învârteau pe un ax. Palele erau albe pe o parte şi negre pe cealaltă. Fața albă a fiecărei pale reflecta lumina, iar fața neagră o absorbea. (De-asta e mai răcoritor să porți un tricou alb şi nu unul negru într-o zi caniculară.) Când asupra dispozitivului se aplica un fascicul de lumină, palele se roteau, fețele negre depărtându-se de lumină. Aceasta este o demonstrație directă că fotonii posedă un impuls suficient de mare pentru a face palele radiometrului să se mişte.[9]

Problema cu care s-a luptat Einstein a fost că impulsul reprezintă o funcție de masă: impulsul este egal cu masa de înmulțit cu viteza. Astfel, o locomotivă care se deplasează cu 48 de kilometri pe oră are un impuls mult mai mare decât, să spunem, o insectă care se deplasează cu aceeaşi viteză. Şi atunci cum poate o particulă cu masa zero să aibă un impuls pozitiv?

Experimentul mintal al lui Einstein a luat în calcul o cutie care pluteşte în spațiu. În interiorul cutiei este lansat un foton, de la stânga la dreapta.



Un radiometru Crookes – girueta cu patru pale se roteşte când asupra ei se proiectează lumină

Impulsul total al sistemului trebuie să se conserve, astfel încât cutia ar fi de aşteptat să reculeze spre stânga la emiterea fotonului. După un anumit interval de timp, fotonul se ciocneşte de faţa din dreapta a cutiei, transferându-i acesteia impulsul înapoi. Impulsul total al sistemului se conservă din nou, astfel încât cutia nu se mai mişcă.

Toate bune până aici. Dar să ne gândim din perspectiva punctului de observaţie al domnului Einstein, care priveşte cutia din afară. Nu vede nicio influenţă exterioară asupra cutiei. Nicio particulă – cu sau fără masă – nu loveşte cutia şi nimic nu iese din ea. Cu toate acestea, domnul Einstein, potrivit scenariului de mai înainte, vede cutia mişcându-se trecător la stânga şi apoi oprindu-se. Potrivit analizei noastre, toţi fotonii ar trebui să deplaseze cutia permanent spre stânga. Cum n-a existat niciun efect exterior asupra cutiei şi nici vreunul generat din cutie, centrul de masă al acesteia trebuie să rămână în acelaşi loc, deoarece fotonul dinăuntrul cutiei, care se deplasează de la stânga la dreapta, nu poate schimba centrul de greutate, deoarece nu are masă.

Sau poate are? Concluzia lui Einstein a fost că, de vreme ce fotonul în mod evident are energie şi impuls, trebuie să aibă şi un echivalent al masei. Energia fotonului aflat în mişcare este echivalentă întru totul cu a unei mase în mişcare. Putem calcula echivalenţa admiţând că centrul de greutate al sistemului trebuie să rămână staţionar în timpul mişcării fotonului. Realizând suportul matematic, Einstein a arătat că masa şi energia sunt echivalente şi sunt legate de o constantă simplă. Pe de altă parte, exista o chichiţă: constanta poate că era simplă, dar s-a dovedit a fi enormă; era viteza luminii la pătrat (în jur de $1,7 \times 10^{17}$ metri pătraţi pe secundă la pătrat – respectiv 17 urmat de 16 zerouri). De aici obţinem celebra formulă a lui Einstein $E = mc^2$.[10] Prin urmare, o uncie (28 de grame) de masă este echivalentă cu 600.000 de tone de trinitrotoluen. Scrisoarea lui Einstein din data de 2 august 1939 adresată preşedintelui Roosevelt, prin care îl informa despre potenţialul unei bombe atomice bazate pe această formulă, a inaugurat epoca atomică.[11]

Ai crede că acest lucru trebuia să fie evident de mai mult timp, dat fiind faptul că experimentatorii observaseră că masa substanţelor radioactive se reducea în timp, ca rezultat al radiaţiei. Dar s-a presupus că substanţele radioactive conţineau un combustibil special de înaltă energie care ardea. Această ipoteză nu este complet greşită; doar că acel combustibil „ars" era, pur şi simplu, masa.

Există mai multe motive pentru care am început această carte cu experimentele mintale ale lui Darwin şi Einstein. În primul rând, ele demonstrează extraordinara forţă a creierului uman. Fără niciun echipament în afara unui creion şi a unei hârtii, pentru desenarea prin linii a personajelor din aceste experimente mintale simple şi pentru scrierea ecuaţiilor cât se poate de simple care rezultă din ele, Einstein a reuşit să răstoarne o înţelegere a lumii fizice care data din urmă cu două secole, să influenţeze profund cursul istoriei (inclusiv al Doilea Război Mondial) şi să inaugureze era nucleară.

Este adevărat, Einstein s-a bazat pe câteva descoperiri experimentale ale secolului al nouăsprezecelea, cu toate că nici în aceste experimente nu s-a folosit vreun echipament special. Este de asemenea adevărat că, în validările experimentale ulterioare ale teoriilor lui Einstein, s-au folosit tehnologii avansate şi că, dacă acestea nu ar fi fost concepute, nu am beneficia de confirmarea pe care o avem astăzi la dispoziţie că ideile lui Einstein sunt autentice şi importante. Pe de altă parte însă, asemenea factori nu umbresc cu nimic faptul că aceste faimoase experimente mintale dezvăluie forţa gândirii umane la nivelul ei cel mai înalt.

Einstein este larg recunoscut drept cel mai însemnat om de ştiinţă al secolului al douăzecilea (şi Darwin ar fi un bun pretendent la această onoare pentru secolul nouăsprezece), deşi suportul matematic care se află la baza teoriilor lui nu este, în ultimă instanţă, deosebit de complicat. Iar experimentul mintal propriu-zis a fost cât se poate de simplu. Şi atunci am putea să ne întrebăm din ce punct de vedere ar trebui să-l considerăm pe Einstein deosebit de deştept? Vom discuta mai târziu despre ce anume a făcut el mai exact în creierul lui până să vină cu aceste teorii şi unde anume sălăşluieşte această calitate.

Pe de altă parte, o asemenea istorie demonstrează şi limitele gândirii omeneşti. Einstein a putut călări pe raza de lumină fără să cadă de pe ea (deşi a concluzionat că era imposibil de fapt să călărească o rază de lumină), însă câte mii de observatori şi gânditori n-au fost absolut lipsiţi de capacitatea de a gândi asemenea exerciţii remarcabil de necomplicate? Un neajuns obişnuit îl reprezintă dificultatea cu care se confruntă majoritatea oamenilor în privinţa eliminării şi depăşirii ideilor şi a perspectivelor semenilor lor. Şi mai există şi alte neajunsuri, pe care le vom discuta în detaliu după ce examinăm felul cum funcţionează neocortexul.

Un model unificat al neocortexului

Cel mai important motiv pentru care îți împărtășesc, cititorule, poate cele mai faimoase experimente mintale din istorie este acela că ele constituie o introducere în utilizarea aceleiași abordări la analizarea creierului. După cum vei vedea, putem ajunge remarcabil de departe în desluşirea modului cum funcționează inteligența umană prin intermediul câtorva simple experimente mintale proprii. Dacă avem în vedere materia implicată, experimentele mintale ar trebui să constituie o abordare cât se poate de potrivită.

Dacă gândurile leneşe ale unui tânăr şi folosirea unui echipament care nu a depăşit un creion şi o hârtie au fost suficiente pentru a revoluționa înțelegerea fizicii, atunci ar trebui să fim în măsură să realizăm un progres remarcabil recurgând la un fenomen cu care suntem mult mai familiarizați. La urma urmelor, cu toții trăim experiența gândirii în fiecare clipă a vieții noastre – în starea de veghe şi, de asemenea, în lumea viselor.

După ce vom construi un model al felului cum se petrec lucrurile prin acest proces de autoreflecție, vom examina măsura în care îl putem confirma prin intermediul celor mai noi observații realizate pe creier real şi prin recrearea, la nivel de ultimă oră, a acestor procese cu ajutorul maşinilor.

Capitolul 2

EXPERIMENTELE MINTALE DE GÂNDIRE

Foarte rar se întâmplă să gândesc în cuvinte. Am câte un gând şi pot încerca să-l exprim prin cuvinte după aceea.

— Albert Einstein

Creierul este o masă de un kilogram şi jumătate pe care o poţi ţine în mână şi care poate să conceapă un univers cu diametrul de o sută de miliarde de ani-lumină.

— Marian Diamond

Ce mi se pare uluitor este faptul că un simplu obiect de un kilogram şi jumătate, compus din aceiaşi atomi care alcătuiesc toate celelalte lucruri de sub Soare, este capabil să conducă absolut tot ce a făcut omul: să zboare pe Lună şi să înscrie şaptezeci de eseuri, să scrie *Hamlet* şi să construiască Taj Mahal-ul – ba chiar şi să dezlege secretele creierului însuşi.

— Joel Havemann

Am început să gândesc asupra gândirii în jurul anului 1960, acelaşi an în care am descoperit computerul. Astăzi ţi-ar fi foarte greu să găseşti un copil de doisprezece ani care să nu folosească un computer, dar pe vremea aceea erau doar câţiva în această situaţie în oraşul meu natal, New York. Desigur, aceste dispozitive timpurii nu puteau fi ţinute în mână şi primul la care am avut acces ocupa o încăpere mare. La începutul anilor 1960, am făcut nişte tentative de programare pe un computer IBM 1620 pentru analize de multiplicitate (un test statistic), cu date care fuseseră adunate prin studierea unui program pentru educaţia la vârste mici – un precursor al programului *Head Start*. Prin urmare, efortul presupunea un dramatism considerabil, soarta acestei iniţiative educative la nivel naţional depinzând de lucrarea noastră. Algoritmii şi datele analizate erau suficient de complexe încât să nu putem anticipa răspunsurile pe care urma să le dea computerul. Bineînţeles, răspunsurile erau determinate de date, dar nu erau predictibile.

S-a dovedit că distincția dintre *determinate* şi *predictibile* este importantă, lucru la care voi reveni.

Îmi amintesc cât de incitant a fost momentul când luminițele de pe panoul frontal s-au estompat, exact înainte ca algoritmul să-şi încheie deliberările, ca şi cum computerul ar fi fost adâncit în gânduri. Când veneau diferite persoane, nerăbdătoare să primească următorul set de rezultate, le arătam cu degetul luminițele care licăreau slab şi le spuneam: „Gândeşte". Era şi nu era o glumă – chiar părea că ar fi cântărit răspunsurile –, iar personalul a început să-i atribuie maşinii o personalitate. Era o antropomorfizare, poate, dar ea m-a făcut să încep să mă gândesc serios la relația dintre gândire şi computare.

Pentru a evalua măsura în care propriul meu creier se asemănă cu programele de computer familiare mie pe atunci, am început să meditez la ce trebuia să facă creierul meu în cursul procesării informației. Am continuat această investigație timp de cincisprezece ani. Ceea ce voi spune în continuare referitor la modul cum înțelegem acum funcționarea creierului îți va părea, dragă cititorule, foarte diferit de conceptul-standard de computer. Pe de altă parte, însă, creierul chiar stochează şi procesează informații şi, datorită universalității computației – concept la care de asemenea mă voi întoarce –, există mai multe paralele între creier şi computer decât ți s-ar părea la prima vedere.

De fiecare dată când fac ceva – sau mă gândesc la ceva –, indiferent că este vorba de periatul dinților, de păşitul prin bucătărie, de contemplarea unei probleme de afaceri, de exersatul la clape sau de găsirea unei noi idei, reflectez asupra felului cum am reuşit s-o obțin. Mă gândesc şi mai mult la toate lucrurile pe care nu sunt capabil să le fac, deoarece limitările gândirii omeneşti asigură un set la fel de important de indicii. Faptul că mă gândesc atât de mult la gândire ar putea foarte bine să mă încetinească, dar am sperat că asemenea exerciții de autoreflecție îmi vor permite să-mi rafinez metodele mintale.

Pentru a ne îmbogăți conştientizarea modului de funcționare a propriului creier, hai să luăm în considerare o serie de experimente mintale.

Încearcă asta: *Spune alfabetul.*

Probabil că îți aminteşti experimentul din copilărie şi îl poți face uşor.

În regulă, acum încearcă asta: *Spune alfabetul de la coadă la cap.*

Dacă nu ai studiat alfabetul în această ordine, este posibil să constați că îți este imposibil să realizezi operația. Ocazional, câte cineva care a

petrecut foarte mult timp într-o sală de clasă de şcoală elementară, unde alfabetul este expus pe pereţi, va putea să facă apel la memoria vizuală şi apoi să-l citească, graţie ei, de la coadă la cap. Însă chiar şi aşa e dificil, deoarece nu ne amintim imagini complete. Recitarea alfabetului în sens invers ar trebui să fie o sarcină simplă, deoarece implică exact aceeaşi informaţie ca şi citirea în sens direct, dar, cu toate acestea, în general suntem incapabili să o facem.

Îţi ţii minte numărul de asigurări sociale? În caz că da, poţi să-l spui invers fără să-l scrii? Dar poezioara de grădiniţă *Mary avea un mieluşel?* Computerele pot face acest lucru în mod curent. Cu toate acestea, noi dăm greş, dacă nu învăţăm în mod expres succesiunea inversă ca pe o serie nouă. Acest fapt ne spune un lucru foarte important despre felul cum este organizată memoria omului.

Desigur, suntem capabili să facem această operaţie cu uşurinţă dacă scriem succesiunea şi apoi o citim de la coadă la cap. Procedând astfel, folosim o tehnologie – limbajul scris – pentru compensarea uneia dintre limitările gândirii noastre neasistate, chit că instrumentul este foarte rudimentar. (A fost cea de-a doua invenţie a noastră, limbajul vorbit fiind prima.) Acesta este motivul pentru care inventăm instrumente – să ne compensăm neajunsurile.

Asta sugerează că **memoria omului este secvenţială şi ordonată. Ea poate fi accesată în ordinea în care e reţinută. Nu suntem capabili să inversăm direct secvenţa unei amintiri.**

De asemenea, avem dificultăţi cu declanşarea unei amintiri din mijlocul unei succesiuni. Dacă învăţ să interpretez o piesă muzicală la pian, în general nu o pot începe dintr-un punct arbitrar de la mijloc. Există câteva puncte la care pot sări, deoarece memorarea secvenţială a piesei este organizată în segmente. Însă, dacă încerc să încep în mijlocul unui segment, trebuie să mă întorc la citirea vizuală până când îmi revine memoria secvenţială.

În continuare, încearcă asta: *Rememorează o plimbare pe care ai făcut-o cam cu o zi înainte. Ce îţi aminteşti din cursul ei?*

Acest experiment mintal merge cel mai bine dacă ai făcut o plimbare recentă, cum ar fi astăzi sau ieri. (Poţi să o înlocuieşti cu o plimbare cu maşina sau, practic, cu orice altă activitate în cursul căreia te-ai deplasat pe un teren.)

Este posibil să nu-ţi aminteşti prea multe despre această experienţă. Care a fost cea de-a cincea persoană pe care ai întâlnit-o (fără să includi

numai oamenii cunoscuți)? Ai văzut un stejar? O cutie poştală? Ce ai văzut după ce ai dat primul colț? Dacă ai trecut pe lângă nişte magazine, ce era în a doua vitrină? Probabil că poți să reconstruieşti răspunsurile la unele dintre aceste întrebări din câteva indicii pe care ți le aminteşti, dar este posibil să-ți reaminteşti doar puține detalii, cu toate că experiența e foarte recentă.

Dacă faci plimbări în mod regulat, gândeşte-te la prima plimbare pe care ai făcut-o luna trecută (sau la primul drum până la birou de luna trecută, dacă faci naveta). Probabil că nu poți să-ți aminteşti deloc respectiva plimbare sau călătorie ca navetist, iar dacă ți-o aminteşti, fără îndoială că îți aduci aminte mai puține detalii legate de ea decât de plimbarea de astăzi.

Voi discuta mai târziu subiectul conştienței şi voi demonstra că tindem să ne adecvăm conştiența în funcție de amintirea evenimentelor. Motivul principal pentru care credem că nu suntem conştienți când ne aflăm sub efectul anesteziei este acela că nu ne amintim nimic din perioada respectivă (cu toate că există excepții ciudate – şi tulburătoare – în această privință). Aşadar, cu privire la plimbarea pe care am făcut-o astăzi de dimineață, n-am fost conştient în cea mai mare parte a ei? Este o întrebare rațională, dat fiind faptul că nu-mi amintesc aproape nimic despre ce am văzut şi la ce m-am gândit.

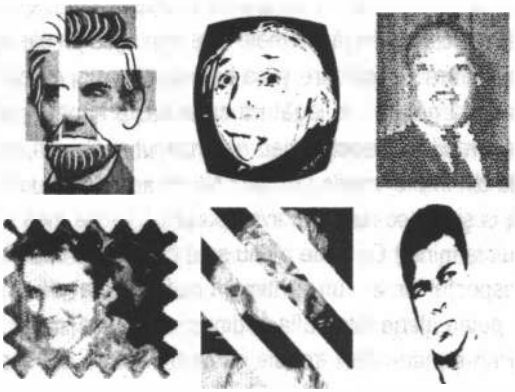
Se întâmplă să existe câteva lucruri pe care mi le amintesc din timpul plimbării de azi dimineață. Îmi aduc aminte că m-am gândit la această carte, dar n-aş putea să-ți spun exact la ce anume m-am gândit. De asemenea, îmi amintesc că am trecut pe lângă o femeie care împingea un cărucior. Îmi aduc aminte că femeia era atrăgătoare şi că bebeluşul era şi el foarte drăgălaş. Îmi amintesc două gânduri în legătură cu această experiență: *Bebeluşul acesta este adorabil, ca nepotul meu nou-născut* şi *Ce o fi înțelegând bebeluşul acesta din împrejurimile vizuale?* Nu-mi amintesc cum era îmbrăcat vreunul dintre ei şi nici ce culoarea avea părul lor. (Soția mea ți-ar spune că e ceva obişnuit la mine.) Cu toate că nu sunt capabil să descriu ceva exact cu privire la aspectul lor, am un sentiment inefabil legat de aspectul femeii şi cred că aş putea alege fotografia ei dintre mai multe imagini cu semene de-ale sale. Prin urmare, deşi trebuie să fie o frântură din aspectul ei cea care mi-a rămas în memorie, dacă mă gândesc la femeie, la cărucior şi la copil, nu sunt capabil să vizualizez ansamblul. Nu există nicio înregistrare fotografică sau video a acestui eveniment în mintea mea. Este greu să descriu exact ce anume din mintea mea e legat de această experiență.

Îmi mai amintesc că am mai trecut pe lângă o femeie cu un cărucior în timpul unei plimbări de acum câteva săptămâni. În acest caz, nu cred că aş mai putea să o recunosc pe femeia aceea într-o fotografie. Amintirea de atunci este mult mai estompată astăzi decât trebuie să fi fost la scurt timp după acea plimbare.

În continuare, gândeşte-te la oamenii pe care i-ai întâlnit doar o dată sau de două ori. Îi poţi vizualiza clar? Dacă eşti artist în domeniul vizual, e posibil să fi învăţat această abilitate, dar de regulă nu suntem capabili să vizualizăm oameni pe care i-am întâlnit doar ocazional, să-i desenăm sau să-i descriem suficient de bine. Nu am avea însă dificultăţi prea mari în recunoaşterea unei fotografii a lor.

Asta sugerează că în creier **nu se stochează imagini, video sau sonore. Amintirile noastre sunt stocate sub forma unor secvenţe de modele. Amintirile care nu sunt accesate se estompează odată cu trecerea timpului.** Când artiştii plastici de la poliţie intervievează victima unei infracţiuni, nu o întreabă *Cum erau sprâncenele infractorului?* În schimb, îi arată victimei o serie de imagini de sprâncene şi îi cer să o aleagă pe una dintre ele. Setul corect de sprâncene va genera recunoaşterea modelului stocat în memoria victimei.

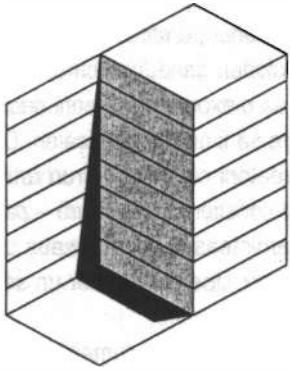
Să ne gândim acum la chipurile pe care le cunoşti bine. *Îl poţi recunoaşte pe vreunul dintre aceşti oameni?*



Fără îndoială că reuşeşti să identifici personaje familiare, cu toate că sunt parţial acoperite sau distorsionate. Acest fapt constituie unul dintre

punctele-cheie ale percepției umane: **putem recunoaşte un model chiar dacă percepem (vedem, auzim, atingem) numai o parte din el şi chiar dacă modelul conține distorsiuni. Puterea noastră de recunoaştere este capabilă, după câte s-ar părea, să detecteze trăsături invariabile ale modelului – caracteristici care supraviețuiesc variațiilor din lumea reală.** Aparentele distorsiuni dintr-o caricatură sau din anumite forme de artă (vezi impresionismul) subliniază modelele unei imagini (persoană, obiect) pe care le recunoaştem, deşi anumite detalii sunt schimbate. Lumea artei se află, de fapt, în avans față de lumea ştiinței în înțelegerea forței sistemului de percepție al omului. Folosim aceeaşi abordare şi atunci când recunoaştem o melodie după numai câteva note.

Acum gândeşte-te la această imagine:



Imaginea este ambiguă – colțul indicat de regiunea neagră poate să fie unul interior sau unul exterior. La început, e posibil să-l percepi într-un fel sau în altul, deşi, cu puțin efort, îți poți modifica percepția mergând către o interpretare alternativă. Însă, din momentul în care mintea ta s-a fixat asupra unui înțeles, îți poate fi foarte dificil să sesizezi cealaltă perspectivă. (Acest lucru se dovedeşte a fi adevărat şi în privința perspectivelor intelectuale.) Interpretarea unei imagini de către creierul tău influențează practic felul în care trăieşti experiența imaginii respective. Când colțul pare a fi unul interior, creierul tău va interpreta regiunea cenuşie ca fiind o umbră; asta înseamnă că zona nu îți va părea atât de întunecată ca atunci când interpretezi colțul ca fiind unul exterior.

Prin urmare, **trăirea conştientă a experienței percepțiilor noastre este modificată, de fapt, de interpretările pe care le dăm experienței respective.**

Consideră că vedem ceea ce ne aşteptăm să_____.

Sunt încrezător că ai reuşit să completezi propoziția de mai înainte.

Dacă aş fi scris ultimul cuvânt, ai fi avut nevoie doar de o ocheadă instantanee pentru a-ți confirma că termenul cu pricina e cel la care te aşteptai.

Asta presupune că **anticipăm viitorul în permanență şi formulăm ipoteze cu privire la experiențele pe care le vom trăi. Aceste aşteptări influențează ceea ce percepem de fapt.** Anticiparea viitorului este, în principal, motivul fundamental pentru care avem un creier.

Să ne gândim la o experiență pe care o trăim cu toții în mod regulat: o amintire din urmă cu ani de zile ne vine în mod inexplicabil în minte.

Adeseori este vorba despre amintirea unei persoane sau a unui eveniment la care nu te-ai mai gândit de mult. Este evident că a existat ceva care a provocat amintirea. Firul gândurilor care a făcut posibil acest lucru poate fi evident şi poate fi unul pe care să-l poți exprima. Alteori, e posibil să fii conştient de succesiunea gândurilor care au dus la revenirea acestei amintiri, dar îți vine greu să o exprimi. Adeseori, declanşatorul dispare rapid, astfel încât amintirea pare să fi apărut de nicăieri. De multe ori mă confrunt cu asemenea amintiri aleatorii când fac lucruri rutiniere, cum ar fi periajul dinților. Uneori pot să fiu conştient de legătură – pasta de dinți ce cade de pe periuță poate să-mi amintească de vopseaua care mi-a căzut de pe o pensulă la un curs din liceu. Uneori am doar un sentiment vag al legăturii sau nu am niciun fel de sentiment.

Un fenomen asociat pe care toată lumea îl trăieşte frecvent este încercarea de a-ți reaminti un nume sau un cuvânt. Procedura pe care o folosim în asemenea împrejurări este să încercăm să vizualizăm declanşatorii capabili să deblocheze amintirea. (De exemplu: *Cine a jucat-o pe Regina Padmé în **Războiul stelelor: Episodul III – Răzbunarea lui Sith**? Să vedem: e aceeaşi actriță care a jucat şi rolul principal într-un film recent despre dans, îi zicea **Lebăda neagră**, a, da, Natalie Portman.*) Uneori utilizăm metode de mnemotehnică idiosincretică pentru a ne reaminti. (De exemplu, *E mereu zveltă, nu cât portu', a, da, Portman, Natalie Portman.*) Unele dintre amintirile noastre sunt suficient de solide încât să ne conducă direct de la o întrebare (ca de exemplu *Cine a jucat rolul Reginei Padmé?*) la răspuns; alteori, trebuie să verificăm o serie de declanşatori pentru a-l găsi pe cel ce se potriveşte. Este o situație care se aseamănă foarte mult cu găsirea link-ului web adecvat. Într-adevăr, e posibil ca amintirile să fie uitate la fel ca o pagină web cu care nu are link-uri nicio altă pagină (cel puțin nicio pagină dintre cele pe care le putem găsi).

În timp ce faci activități de rutină – cum ar fi îmbrăcarea cămășii –, urmărește-te efectuându-le și gândește-te la măsura în care urmezi exact aceeași succesiune a pașilor de fiecare dată. Din propriile mele observații (și, cum am menționat, încerc să mă observ permanent), deduc că este posibil să urmezi în cea mai mare măsură aceiași pași de fiecare dată când îndeplinești o anumită activitate de rutină, cu toate că pot să intervină și alte module, adăugate. De exemplu, cele mai multe dintre cămășile mele nu au nevoie de butoni, dar, când e una care are, asta implică o altă serie de sarcini.

Lista pașilor din mintea mea este organizată pe ierarhii. Îndeplinesc o anumită rutină înainte să mă duc la culcare. Primul pas este să-mi periez dinții. Dar această acțiune este la rândul ei descompusă în serii de pași mai mici, dintre care primul e să pun pastă pe periuță. Acest pas, la rândul lui, e alcătuit din pași și mai mici, cum ar fi găsirea tubului de pastă, scoaterea capacului și așa mai departe. Pasul găsirii tubului are și el pașii lui, primul dintre aceștia fiind deschiderea dulapului din baie. Acest pas, la rândul său, necesită alți pași, primul fiind apucarea exteriorului de la ușa dulapului. Această ramificare continuă, practic, până la o serie de mișcări foarte fine, astfel încât există literalmente mii de acțiuni mici care formează rutina mea de culcare. Cu toate că mă confrunt poate cu dificultăți de reamintire a detaliilor plimbării pe care am făcut-o cu numai câteva ore în urmă, nu am nicio greutate în a-mi aminti toți aceștia nenumărați pași ai pregătirii de culcare – într-o asemenea măsură încât sunt în stare să mă gândesc și la alte lucruri în timp ce parcurg aceste proceduri. Este important să scot în evidență faptul că această listă nu este stocată ca o lungă listă cu mii de pași – **dimpotrivă, fiecare dintre procedurile noastre rutiniere e reținută sub forma unei ierarhii elaborate de activități asociate.**

Același tip de ierarhie este implicat și în capacitatea noastră de recunoaștere a obiectelor și situațiilor. Recunoaștem chipurile oamenilor pe care îi cunoaștem bine și recunoaștem de asemenea faptul că aceste chipuri au ochi, un nas, o gură și așa mai departe – o ierarhie de modele pe care o folosim atât în percepțiile, cât și în acțiunile noastre. Utilizarea ierarhiilor ne permite să refolosim modelele. De exemplu, nu trebuie să reînvățăm conceptul de nas și gură de fiecare dată când facem cunoștință cu un chip nou.

În următorul capitol, vom reuni rezultatele acestor experimente mintale într-o teorie a modului în care ar trebui să funcționeze neocortexul. Voi demonstra că experimentele dezvăluie atribute esențiale ale gândirii noastre care sunt uniforme, de la găsirea pastei de dinți, până la compunerea unei poezii.

Capitolul 3

UN MODEL DE NEOCORTEX: TEORIA MINȚII BAZATE PE RECUNOAȘTERE

Creierul este un țesut. E un țesut complicat, minuțios urzit, cum nu mai e nimic altceva din ce cunoaștem în Univers, dar este format din celule, ca orice țesut. În mod evident, sunt celule superior specializate, dar ele funcționează potrivit legilor care guvernează toate celelalte celule. Semnalele lor electrice şi chimice pot fi detectate, înregistrate şi interpretate, iar compuşii lor chimici pot fi identificați; conexiunile care constituie textura creierului pot fi cartografiate. Pe scurt, creierul poate fi studiat exact aşa cum poate fi studiat şi rinichiul.

— David H. Hubel, specialist în neuroştiință

Să presupunem că ar exista o maşină a cărei structură ar produce gândirea, senzațiile şi percepțiile; să ne imaginăm această maşină mărită dar păstrând aceleaşi proporții, astfel încât să poți intra în ea ca şi cum ar fi o fabrică. Cu aceste ipoteze, i-ai putea vizita interiorul. Şi ce-ai putea observa acolo? Nimic altceva decât componente care se împing şi se mişcă unele pe altele, niciodată ceva ce ar putea să explice percepția.

— Gottfried Wilhelm Leibniz

O ierarhie a modelelor

Am repetat experimentele şi observațiile simple descrise în capitolul anterior de mii de ori, într-o multitudine de contexte. Concluziile acestor observații mă obligă să explic ce trebuie să facă creierul, la fel cum simplele experimente cu timpul, spațiul şi masa care s-au efectuat la începutul şi sfârşitul secolului al nouăsprezecelea au impus reflecțiile tânărului maestru Einstein cu privire la funcționarea Universului. În cadrul discuției ce urmează, voi include şi câteva observații elementare din domeniul neuroştiinței, încercând să evit numeroasele detalii care rămân în continuare controversate.

În primul rând, dă-mi voie să explic de ce această secțiune discută exact neocortexul (din latină, cu sensul de *scoarță nouă*). Știm cu siguranță că neocortexul este răspunzător de capacitatea noastră de a face față modelelor de informație și că face acest lucru de o manieră ierarhică. Animalele lipsite de neocortex (în esență nemamiferele) sunt în mare parte incapabile de înțelegerea ierarhiilor.[1] Înțelegerea și supunerea față de natura ierarhică înnăscută a realității este o caracteristică exclusivă a mamiferelor și rezultă din faptul că mamiferele sunt singurele care posedă această recentă structură evolutivă a creierului. Neocortexul este răspunzător de percepția senzorială, de recunoașterea a tot ce există, de la obiectele vizuale la conceptele abstracte, la controlul mișcărilor, la judecată, de la orientarea spațială la gândirea rațională și la limbaj – adică, în esență, ceea ce considerăm că reprezintă „gândirea".

Neocortexul omului, stratul exterior al creierului, este o structură subțire, în esență bidimensională, cu grosimea de aproximativ 2,5 milimetri (în jur de o zecime de inch). La rozătoare are aproximativ dimensiunile unui timbru poștal și este neted. O inovație evolutivă la primate este aceea că neocortexul s-a cutat minuțios peste restul creierului, având adâncituri profunde, șanțuri și încrețituri care-i măresc suprafața. Datorită acestei configurații elaborate, neocortexul reprezintă majoritatea creierului omului, revenindu-i 80 la sută din greutatea acestuia. *Homo sapiens* a căpătat o frunte înaltă pentru a face posibil un neocortex și mai mare; avem, în special, un lob frontal care se ocupă de cele mai abstracte modele asociate cu conceptele de nivel înalt.

Această structură subțire este compusă, în esență, din șase straturi, numerotate de la I (cel exterior) la VI. Axonii neuronilor din straturile II și III se proiectează și în alte regiuni ale cortexului. Axonii (conexiunile de ieșire) din straturile V și VI sunt conectați în principal în afara neocortexului, cu talamusul, trunchiul cerebral și măduva spinării. Neuronii din stratul IV primesc conexiuni sinaptice (de intrare) de la neuroni care se află în afara neocortexului, în special de la talamus. Numărul de straturi variază puțin de la o regiune la alta. Stratul IV este foarte subțire în neocortexul motor, deoarece, în această regiune, în cea mai mare parte nu primește intrări de la talamus, de la trunchiul cerebral sau de la măduva spinării. În schimb, în lobul occipital (partea neocortexului răspunzătoare în general de procesarea vizuală) există trei straturi suplimentare care pot fi văzute în stratul IV, datorită intrărilor considerabile care vin în această regiune, inclusiv de la talamus.

O observație de o importanță fundamentală legată de neocortex privește extraordinara uniformitate a structurii sale fundamentale. Aceasta a fost observată pentru prima oară de către un specialist în neuroştiință, americanul Vernon Mountcastle (născut în anul 1918). În 1957, Mountcastle a descoperit organizarea neocortexului în formă de coloane. În anul 1978, el a făcut o observație la fel de semnificativă pentru neuroştiință ca experimentul Michelson-Morley – care, în anul 1887, a infirmat existența eterului – pentru fizică. În acel an 1978, Mountcastle a descris organizarea invariabilă a neocortexului, emițând ipoteza că acesta este format dintr-un singur mecanism care se repetă de foarte multe ori[2] şi propunând ca unitate de bază *coloana corticală*. Diferențele de înălțime dintre anumite straturi din regiunile menționate anterior sunt pur şi simplu diferențe ale volumului de interconectare căruia regiunile respective sunt răspunzătoare să îi facă față.

Mountcastle a emis şi ipoteza existenței unor minicoloane în interiorul coloanelor, însă această teorie a fost controversată, din cauză că nu existau demarcații vizibile între aceste structuri mult mai mici. Însă experimentele ulterioare au dezvăluit că există, de fapt, unități care se repetă în cadrul texturii neuronale a fiecărei coloane. Părerea mea este că unitatea de bază e un recognizator de modele şi că acesta constituie componenta fundamentală a neocortexului. Contrar concepției lui Mountcastle referitoare la minicoloane, nu există limite specifice între aceşti recognizori, ei fiind plasați unul lângă celălalt de o manieră întrepătrunsă, astfel încât coloana corticală reprezintă pur şi simplu un agregat compus dintr-un mare număr de recognizatori. Aceşti recognizori sunt capabili să se interconecteze între ei pe tot parcursul vieții. Prin urmare, interconectarea complexă (dintre module) pe care o observăm la nivelul neocortexului nu este prespecificată de codul genetic, ci e creată pentru a reflecta modelele pe care le învățăm propriu-zis de-a lungul vieții. Voi prezenta mai detaliat teza enunțată, dar susțin că acesta e modul în care probabil că este organizat neocortexul.

Ar trebui precizat, înainte să analizăm mai departe structura neocortexului, că este important să se realizeze modele ale sistemelor la nivelul adecvat. Deşi chimia se bazează teoretic pe fizică şi poate fi dedusă pe de-a-ntregul din fizică, acest lucru ar fi lipsit de rezultate şi nefezabil în realitate, deoarece chimia şi-a stabilit propriile reguli şi modele. În mod similar, ar trebui să putem deduce din fizică legile termodinamicii, dar, din momentul în care avem un număr suficient de mare de particule ca să le putem numi gaz şi nu nişte particule oarecare, aplicarea ecuațiilor pentru fizica interacțiunilor

fiecărei particule devine imposibil de realizat, în timp ce legile termodinamicii se aplică foarte bine.

La fel, și biologia își are propriile reguli și modele. O singură celulă a unei insule pancreatice este enorm de complicată, mai ales dacă îi creăm modelul la nivelul moleculelor; realizarea unui model pentru ceea ce face de fapt pancreasul sub raportul reglării nivelului sanguin al insulinei și sub cel al secreției enzimelor digestive este considerabil mai puțin complexă.

Același principiu se aplică și nivelurilor de realizare a modelelor și de înțelegere la nivelul creierului. Fără îndoială, realizarea modelului interacțiunilor la nivel molecular este o parte utilă și necesară a procesului de retroinginerie a creierului, dar scopul efortului în acest moment este rafinarea modelului pentru a explica felul în care creierul procesează informația ca să genereze înțelesuri cognitive.

Omul de știință american Herbert A. Simon (1916-2001), care este creditat drept cofondator al domeniului inteligenței artificiale, a scris elocvent pe tema înțelegerii sistemelor complexe la nivelul de abstracție adecvat. În descrierea programului de inteligență artificială conceput de el și numit EPAM (perceptor și memorizator elementar), Simon a scris, în anul 1973: „Să presupunem că vrei să înțelegi misteriosul program EPAM pe care îl dețin. Aș putea să-ți dau două versiuni ale lui. Una ar putea să fie... forma în care l-am scris practic – cu întreaga structură de rutine și subrutine... Pe de altă parte, aș putea să-ți dau o versiune în limbaj-mașină a EPAM după ce s-a realizat traducerea completă – după ce a fost uniformizat, ca să spun așa... Nu cred că trebuie să demonstrez detaliat care dintre aceste două versiuni ar constitui cea mai parcimonioasă, cea mai plină de semnificații, cea mai legitimă descriere... Nici măcar nu ți-o voi propune pe a treia: să nu-ți dau niciun program, dar în schimb să-ți pun la dispoziție ecuațiile electromagnetice și condițiile-limită pe care computerul, considerat sistem fizic, ar trebui să le respecte comportându-se ca EPAM. Acesta ar fi apogeul reducerii și al incomprehensibilului".[3]

Există în jur de o jumătate de milion de coloane corticale într-un creier uman, fiecare ocupând un spațiu de aproximativ doi milimetri înălțime și o jumătate de milimetru diametru și conținând în jur de 60.000 de neuroni (fapt ce se concretizează într-un total de aproximativ 30 de miliarde de neuroni în neocortex). O estimare grosieră spune că fiecare recognizor de modele dintr-o coloană corticală conține aproximativ 100 de neuroni, așa că, în total, în neocortex există ceva de ordinul a 300 de milioane de recognizatori de modele.

Deoarece analizăm modul de funcționare a acestor recognizatori de modele, dă-mi voie să încep prin a spune că este dificil să știu cu precizie de unde să încep. Totul în neocortex se întâmplă simultan, deci nu există nici un început și nici un sfârșit al proceselor sale. Voi fi obligat adeseori să mă refer la fenomene pe care nu le-am explicat încă, dar la care intenționez să mă întorc, astfel încât te rog, cititorule, suportă aceste referințe ulterioare.

Ființa umană are o capacitate redusă de procesare logică, dar posedă o capacitate foarte profundă de recunoaștere a modelelor. Pentru gândirea logică, trebuie să ne folosim neocortexul, care este, în esență, un imens recognizator de modele. Nu e un mecanism ideal pentru realizarea transformărilor logice, însă este singurul mijloc pe care îl avem pentru realizarea acestui scop. Să comparăm, de exemplu, modul în care joacă șah un om și modul de funcționare a unui program tipic de șah pentru computer. Deep Blue, computerul care l-a învins, în anul 1997, pe Garry Kasparov, campionul mondial uman la șah, era capabil să analizeze implicațiile logice pentru un număr de 200 de milioane de poziții pe tabla de joc (reprezentând diferite succesiuni mutare-contramutare) în fiecare secundă. (Apropo, acest lucru poate fi realizat acum de câteva computere personale.) Kasparov a fost întrebat câte poziții putea să analizeze pe secundă, iar el a răspuns că mai puțin de una. Și atunci cum de a fost posibil să reușească să-i facă totuși față lui Deep Blue? Răspunsul constă în capacitatea foarte mare a omului de a recunoaște modele. Pe de altă parte, trebuie să ne antrenăm această capacitate, motiv pentru care nu oricine poate juca șah la nivel de maestru.

Kasparov învățase în jur de 100.000 de poziții pe tabla de joc. Este un număr real – am stabilit că un maestru uman dintr-un anumit domeniu stăpânește în jur de 100.000 de calupuri de informație. Shakespeare și-a compus piesele de teatru cu 100.000 de sensuri ale cuvintelor (folosind în jur de 29.000 de cuvinte diferite, însă aplicându-le în numeroase moduri). Sistemele medicale expert care au fost construite pentru a reprezenta cunoștințele unui medic uman au arătat că un specialist uman din domeniul medical stăpânește în jur de 100.000 de concepte din domeniul său. Recunoașterea unui calup de informație din acest depozit nu este simplă, unul și același element prezentându-se puțin diferit de fiecare dată când este activat.

Înarmat cu aceste cunoștințe, Kasparov privește tabla de șah, compară modelele pe care le vede cu toate cele 100.000 de situații pe care le stăpânește și face toate aceste 100.000 de comparații simultan. Asupra acestui

aspect există un consens. Toți neuronii noștri procesează – analizează – modelele în același timp. Acest lucru nu înseamnă că toți *emit impulsuri* simultan (probabil că am cădea la pământ dacă s-ar întâmpla așa ceva), însă, în timp ce procesează, iau în calcul posibilitatea de a emite impulsuri.

Câte modele poate să stocheze neocortexul? Trebuie să luăm în calcul fenomenul redundanței. Chipul unei persoane dragi, de exemplu, nu este stocat o singură dată, ci de mii de ori. Unele dintre aceste repetiții reproduc în mare măsură aceeași imagine a chipului, dar majoritatea prezintă diverse perspective ale acestuia, în lumini diferite, cu expresii diferite și așa mai departe. Niciunul dintre aceste modele repetate nu este stocat ca imagine propriu-zisă (adică sub forma unor dispuneri bidimensionale de pixeli). Dimpotrivă, e stocat sub forma unor liste de trăsături ale căror elemente proprii modelului sunt ele însele modele. Vom spune în continuare mai exact cum se prezintă aceste ierarhii de caracteristici și cum sunt ele organizate.

Dacă pornim de la premisa că toate cunoștințele fundamentale ale unui expert rezidă în aproximativ 100.000 de calupuri de informație (respectiv modele) cu o redundanță estimată de aproximativ 100 la 1, obținem un necesar de 10 milioane de modele. Aceste cunoștințe fundamentale la nivel de expert sunt edificate pe cunoștințe profesionale mai extinse, deci putem ridica ordinul de mărime al modelelor la circa 30 până la 50 de milioane. Cunoștințele noastre cotidiene „de bun-simț" pentru ființa umană sunt și mai numeroase; „descurcăreții" au nevoie practic de o parte mult mai mare din neocortex decât „erudiții". Incluzând și aceste aspecte, estimarea noastră ajunge la peste 100 de milioane de modele, cu un factor de redundanță de circa 100. A se observa că factorul de redundanță nu este nici pe departe fix – modelele foarte obișnuite înregistrează ușor un factor de redundanță de ordinul miilor, în vreme ce un fenomen nou-nouț poate să aibă factorul de redundanță mai mic de 10.

După cum voi arăta în continuare, procedurile și acțiunile noastre conțin de asemenea modele și sunt stocate tot în regiuni ale neocortexului, astfel încât estimarea pe care o dau eu pentru capacitatea neocortexului uman este de ordinul a câteva sute de milioane de modele. Această evaluare brută se corelează bine cu numărul de recognizatori de modele, pe care îl estimez undeva în jur de 300 de milioane, deci concluzia rezonabilă este că funcția fiecărui recognizator neocortical de modele constă în procesarea unei iterări (cu alte cuvinte, o copie dintre multiplele copii redundante ale majorității modelelor din neocortex) a unui model. Estimarea noastră privitoare la numărul

de modele cu care este capabil să opereze creierul omului (inclusiv redundanța necesară) și numărul de recognizatori fizici de modele se întâmplă să fie de același ordin de mărime. Trebuie scos în evidență aici faptul că, atunci când mă refer la „procesarea" unui model, vorbesc despre toate lucrurile pe care suntem capabili să le facem cu un model: să-l citim, să-l anticipăm (inclusiv părți din el), să-l recunoaștem și să-l implementăm (gândindu-ne mai departe la el sau utilizând un model de mișcare fizică).

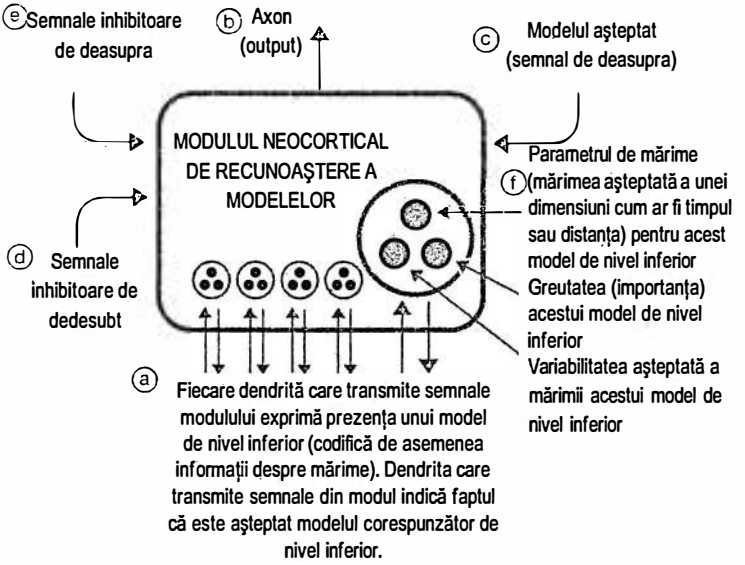
Trei sute de milioane de procesoare de modele poate să pară un număr mare și într-adevăr ar fi suficient pentru a-i permite lui *Homo sapiens* să pună la punct limbajul vorbit și scris, toate uneltele și alte creații. Aceste invenții s-au dezvoltat unele după altele, dând naștere unei creșteri exponențiale a conținutului de informație al tehnologiilor, potrivit descrierii mele din Legea accelerării veniturilor. Nicio altă specie nu a mai realizat acest lucru. După cum am discutat, câteva specii, cum ar fi cimpanzeii, par să posede și ele o abilitate rudimentară de a înțelege și a compune limbajul, precum și unelte primitive. Și ele au, la urma urmelor, un neocortex, însă capacitățile lor sunt limitate din cauza dimensiunilor mai mici ale acestuia, în special la nivelul lobilor frontali. Mărimea neocortexului nostru a depășit un prag care a permis speciei noastre să construiască unelte din ce în ce mai eficiente, inclusiv unele care ne pot permite acum să ne înțelegem propria inteligență. În ultimă instanță, creierul nostru, combinat cu tehnologiile pe care le-a conceput, ne va permite să creăm un neocortex sintetic care va conține mult peste 300 de milioane de procesatori de modele. De ce nu un miliard? Sau o mie de miliarde?

Structura unui model

Teoria minții bazate pe recunoașterea modelelor pe care o voi prezenta aici este întemeiată pe recunoașterea modelelor de către modulele de recunoaștere a modelelor din neocortex. Aceste modele (ca și modulele) sunt organizate în ierarhii. Voi prezenta în cele ce urmează rădăcinile intelectuale ale acestei idei, inclusiv propriile mele lucrări în domeniul recunoașterii ierarhizate a modelelor din anii 1980 și 1990 și modelul de neocortex al lui Jeff Hawkins (născut în 1957) și Dileep George (născut în 1977), de la începutul anilor 2000.

Fiecare model (care este recunoscut de unul dintre cele 300 de milioane de recognizatori de modele estimate că există la nivelul neocortexului)

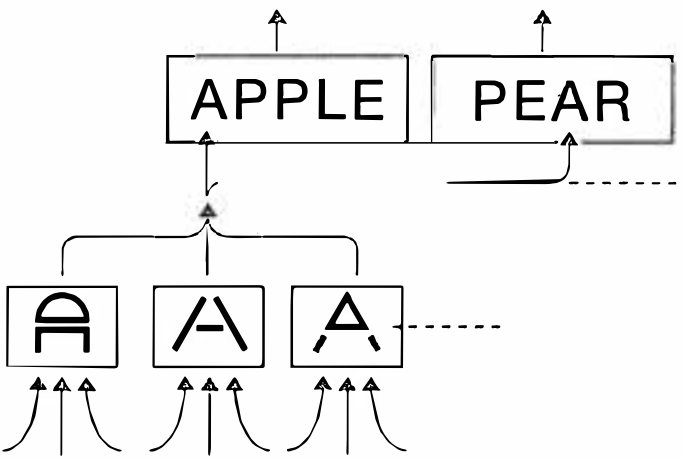
este compus din trei părţi. Prima parte este input-ul, care constă în modelele de nivel inferior ce compun modelul principal. Descrierile fiecăruia dintre aceste modele de nivel inferior trebuie repetată. De exemplu, multe


e Semnalele inhibitoare de deasupra
b Axon (output)
c Modelul aşteptat (semnal de deasupra)
MODULUL NEOCORTICAL DE RECUNOAŞTERE A MODELELOR
Parametrul de mărime
f (mărimea aşteptată a unei dimensiuni cum ar fi timpul sau distanţa) pentru acest model de nivel inferior
d Semnale inhibitoare de dedesubt
Greutatea (importanţa) acestui model de nivel inferior
Variabilitatea aşteptată a mărimii acestui model de nivel inferior
a
Fiecare dendrită care transmite semnale modulului exprimă prezenţa unui model de nivel inferior (codifică de asemenea informaţii despre mărime). Dendrita care transmite semnale din modul indică faptul că este aşteptat modelul corespunzător de nivel inferior.

dintre modelele cuvintelor includ litera „A". Niciunul dintre ele nu are nevoie de repetarea descrierii literei „A", deoarece va folosi aceeaşi descriere. Gândeşte-te la acest lucru ca la un indicator web. Există o pagină web (respectiv un model) pentru litera „A" şi toate paginile web (modelele) pentru cuvintele care conţin „A" vor avea un link la pagina „A" (la modelul „A"). În locul link-urilor pe web, neocortexul foloseşte conexiuni neuronale reale. Există un axon din recognizatorul de modele „A" care se conectează cu mai multe dendrite, câte una pentru fiecare cuvânt care foloseşte „A". Reţine şi factorul de redundanţă: nu există un singur recognizator de modele pentru litera „A". Oricare dintre aceşti recognizatori de modele pentru „A" poate să trimită un semnal la recognizatorii de modele care conţin „A".

A doua parte a fiecărui model este numele acestuia. În lumea limbajului, un model de nivel superior este pur şi simplu cuvântul „apple" („măr"). Cu toate că ne folosim direct neocortexul pentru a înţelege şi procesa

fiecare nivel al limbajului, majoritatea modelelor pe care le conține acesta nu sunt modele de limbaj ca atare. În neocortex, „numele" unui model e pur şi simplu axonul care iese din fiecare procesor de modele; când acest axon emite un impuls, înseamnă că modelul corespunzător a fost recunoscut. Emiterea impulsului de către axon semnifică faptul că recognizatorul strigă numele modelului: „Hei, omule, tocmai am văzut scris cuvântul «apple» («măr»)".

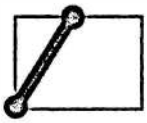


APPLE (măr) PEAR (pară)
Trei modele redundante (dar oarecum diferite) pentru „A" transmise modelelor superioare care conțin „A"

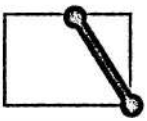
A treia şi ultima parte a fiecărui tipar este setul de modele de nivel superior din care face şi el parte. Pentru litera „A", asta înseamnă toate cuvintele care conțin „A". Acestea sunt, iarăşi, asemenea link-urilor web. Fiecare model recunoscut la un anumit nivel sesizează următorul nivel că aici e prezentă o parte din modelul de nivel superior. La nivelul neocortexului, aceste link-uri sunt reprezentate de dendrite fizice care se duc la neuronii din fiecare recognizator cortical de modele. Reține că fiecare neuron poate primi imput-uri de la mai multe dendrite, dar că generează un singur output printr-un axon. Acest axon însă poate să transmită la rândul lui semnalul către mai multe dendrite.

Să luăm câteva exemple simple: modelele de pe paginile care urmează reprezintă un mic subset de modele folosite pentru alcătuirea literelor tipărite. Se observă că fiecare nivel reprezintă un model. În acest caz, formele sunt modele, literele sunt modele şi cuvintele sunt de asemenea modele. Fiecare dintre aceste modele are un set de input-uri, un proces al recunoaşterii modelului (bazat pe input-uri care se petrec în modul) şi un output (care este transmis următorului nivel superior de recognizatori de modele).

Conexiunea din sud-vest către centru-nord:



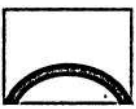
Conexiunea sud-est către centru-nord:



Bara orizontală:

Bară verticală de extremă stângă:

Regiunea concavă cu fața spre sud:



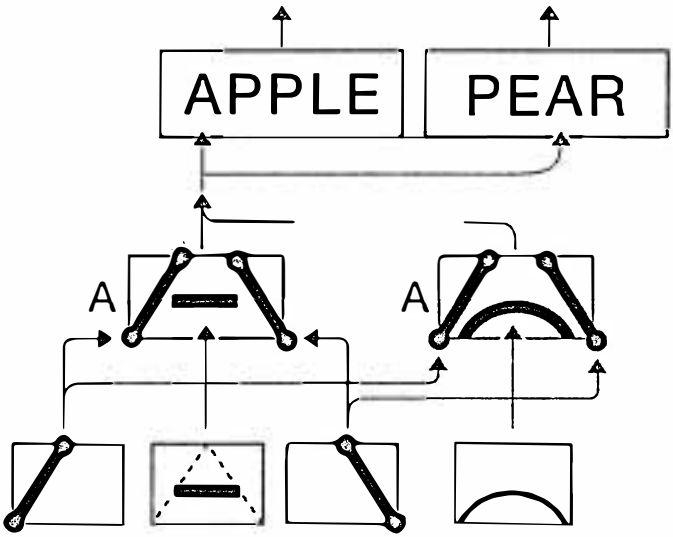
Linia orizontală inferioară:

Linia orizontală superioară:

Linia orizontală medie:

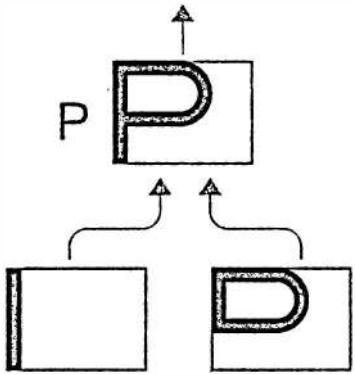
Buclă care formează regiunea superioară:

Modelele de mai înainte sunt componente ale următorului nivel superior – o categorie denumită litere tipărite (în orice caz, la nivelul neocortexului nu există o asemenea categorie formală; şi, de fapt, nu există categorii formale) „A":



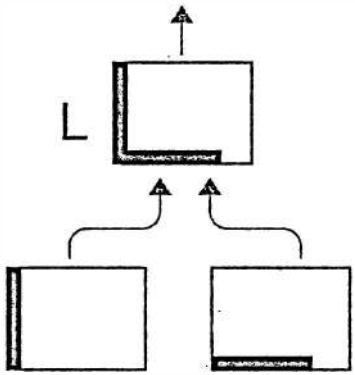
Două modele diferite care constituie fiecare un „A" şi două modele diferite la un nivel superior („APPLE" şi „PEAR" – „măr" şi „pară") din care face parte „A"

„P":



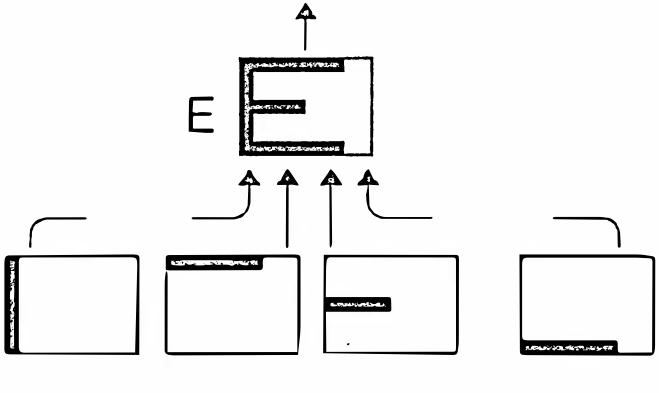
Modele care fac parte din modelul de nivel superior „P"

„L":



Modele care fac parte din modelul de nivel superior „L"

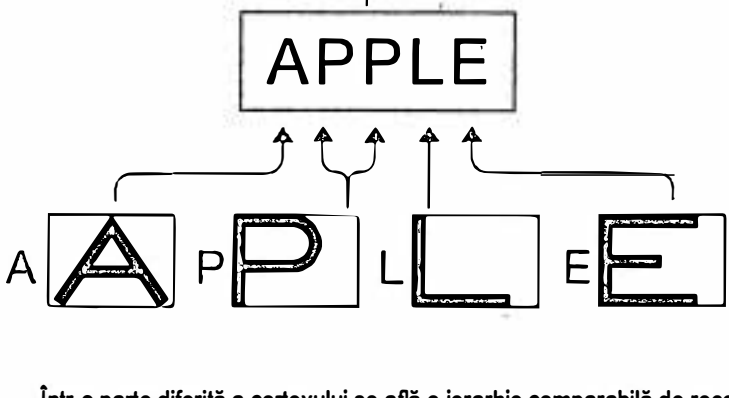
„E":



Modele care fac parte din modelul de nivel superior „E"

Modelele literelor transmise la un model de nivel superior dintr-o categorie numită „cuvinte". (Termenul „cuvinte" este categoria noastră de limbaj pentru acest concept, dar neocortexul îl tratează doar ca model.)

„APPLE":



Într-o parte diferită a cortexului se află o ierarhie comparabilă de recognizatori de modele care procesează imaginile propriu-zise ale obiectelor (şi nu literele tipărite). Dacă priveşti un măr real, recognizatorul de nivel redus detectează marginile curbe şi modelele de culori ale suprafeţei. Acestea fac

ca recognizatorul să emită un impuls prin axon şi să spună, de fapt, „Hei, omule, am văzut un măr real". Alți recognizatori de modele vor detecta combinații de frecvențe ale sunetelor ce ajung la ei din cortexul auditiv, care poate să-şi excite propriul axon indicând: „Am auzit că s-a rostit cuvântul «măr»".

Nu uita factorul redundant – nu avem un singur recognizator de model pentru „măr" în fiecare dintre formele lui (scrisă, vorbită, vizuală). Este posibil să existe sute de asemenea recognizatori care să acționeze, dacă nu chiar mai mulți. Redundanța nu numai că măreşte posibilitatea de a recunoaşte cu succes fiecare dintre aceste ipostaze ale mărului, ci se ocupă şi de variațiile merelor din lumea reală. Pentru merele-obiecte există recognizatori de modele care se ocupă de diversele aspecte ale merelor: culori, nuanțe, forme şi varietăți.

De asemenea, reține că ierarhia prezentată anterior este o ierarhie de concepte. Aceşti recognizatori nu sunt plasați fizic unii deasupra altora; din cauza subțirimii sale, neocortexul are, din punct de vedere fizic, înălțimea egală cu a unui singur recognizator. Ierarhia conceptuală este creată de conexiunile dintre recognizatorii individuali de modele.

Un atribut important al Teoriei minții bazate pe recunoaşterea modelelor îl reprezintă felul în care se fac recunoaşterile în interiorul fiecărui modul de recunoaştere de modele. În modul este stocată o greutate pentru fiecare dendrită de input care arată cât de important este input-ul respectiv pentru recunoaştere. Recognizatorul de modele are un prag de declanşare a impulsului care indică faptul că acest recognizator a recunoscut cu succes modelul pentru care este răspunzător. Nu trebuie să fie prezente toate modelele de input pentru ca recognizatorul să emită impulsul. El poate emite respectivul impuls şi dacă lipseşte un input cu greutate redusă, dar e mai puțin probabil să o facă dacă lipseşte un input cu greutate mare. Când emite impulsul, recognizatorul spune în esență: „Modelul de care sunt răspunzător este probabil prezent".

Recunoaşterea cu succes de către un modul a modelului său depăşeşte simpla numărare a semnalelor de input care sunt activate (chiar şi o numărare căreia îi conferă greutate importanța parametrului). Mărimea fiecărui input contează. Mai există un parametru (pentru fiecare input) care indică mărimea aşteptată a input-ului – şi încă unul care arată cât de variabilă este această mărime. Pentru a înțelege cum funcționează, să presupunem că avem un recognizator de modele care este răspunzător de

recunoaşterea cuvântului „steep" – „abrupt". Acest cuvânt rostit are patru sunete: [s], [t], [E] şi [p]. Fonemul [t] este cunoscut sub numele de „consoană dentală", ceea ce înseamnă că e creat de limbă, care produce o explozie zgomotoasă când aerul scapă din contactul cu dinţii superiori. În esenţă, este imposibil ca fonemul [t] să se articuleze încet. Fonemul [p] este considerat „consoană plozivă" sau „oral ocluzivă", ceea ce înseamnă că e produs atunci când tractul vocal este blocat brusc (de buze în cazul lui [p], astfel încât aerul nu mai trece). Este, de asemenea, obligatoriu, scurt. Vocala [E] este produsă de rezonanţa coardelor vocale cu gura deschisă. E considerată o „vocală lungă", ceea ce înseamnă că persistă o perioadă de timp mult mai lungă decât consoane cum ar fi [t] şi [p]; pe de altă parte, durata poate fi destul de variabilă. Fonemul [s] este cunoscut drept „consoană şuierată" şi e produs de trecerea aerului peste muchiile dinţilor ţinuţi apropiaţi. Durata sa este de obicei mai scurtă decât a unei vocale lungi cum ar fi [E], dar e tot variabilă (cu alte cuvinte, [s] poate fi rostit scurt sau poate fi rostit prelung).

În cercetările noastre dedicate recunoaşterii vorbirii, am descoperit că este necesar ca acest gen de informaţie să fie codificată, pentru recunoaşterea modelelor de vorbire. De exemplu, cuvintele „step" („pas", „treaptă") şi „steep" sunt foarte asemănătoare. Cu toate că fonemul [e] din „step" şi fonemul [E] din „steep" sunt vocale oarecum diferite (prin faptul că au frecvenţe de rezonanţă diferite), nu este de încredere să se facă distincţia între aceste două cuvinte pe baza acestor vocale adeseori confundate între ele. E mult mai de încredere să se ia în considerare observaţia că fonemul [e] din cuvântul „step" este relativ scurt în comparaţie cu fonemul [E] din cuvântul „steep".

Putem codifica acest tip de informaţie cu două cifre pentru fiecare input: mărimea aşteptată şi gradul de variabilitate a acestei mărimi. În exemplul cu „steep", [t] şi [p] au ambele o durată aşteptată foarte scurtă şi o variabilitate aşteptată redusă (adică nu ne aşteptăm să auzim t-uri şi p-uri lungi). Sunetul [s] are o durată aşteptată mai mică, dar o variabilitate mai mare, deoarece poate fi prelungit. Sunetul [E] are atât o durată aşteptată mare, cât şi un grad ridicat de variabilitate.

În exemplele noastre de vorbire, parametrul „mărime" se referă la durată, dar timpul este doar una dintre dimensiunile posibile. De-a lungul cercetărilor noastre de recunoaştere a caracterelor, am constatat că informaţia spaţială comparabilă era mai importantă pentru recunoaşterea literelor tipărite (de exemplu, punctul de pe litera „i" este aşteptat să fie mult mai mic

decât porțiunea de literă de sub punct). La niveluri mult mai înalte de abstractizare, neocortexul poate opera cu modele cu tot felul de continuumuri, cum ar fi nivelurile de atractivitate, ironie, fericire, frustrare şi multe altele. Putem găsi similitudini între continuumuri cât se poate de diferite, cum a făcut Darwin atunci când a asociat mărimea fizică a canioanelor geologice cu măsura diferențierii dintre specii.

În cazul unui creier biologic, sursa acestor parametri vine din experiența proprie a creierului. Nu ne naştem cunoscători ai fonemelor; şi, pe bună dreptate, diferitele limbi au seturi diferite de foneme. Asta presupune faptul că în parametrii învățați ai fiecărui recognizator de modele sunt codificate numeroase exemple ale unui model (deoarece recognizatorul impune mai multe exemple de model pentru a avea siguranța distribuției magnitudinilor input-urilor modelului). În unele sisteme de inteligență artificială, aceste tipuri de parametri sunt codificate manual de către experți (de exemplu de lingviști, care ne pot preciza duratele aşteptate ale diferitelor foneme, după cum am arătat mai înainte). În cercetările mele, am descoperit că o abordare superioară ar presupune existența unui sistem de inteligență artificială care să identifice aceşti parametri singur, din datele achiziționate pe parcursul pregătirii (similar modului în care procedează creierul). Uneori am folosit o abordare hibridă, adică am îmbogățit sistemul cu intuiția experților umani (pentru setările inițiale ale parametrilor), iar pe urmă am determinat sistemul de inteligență artificială să-şi rafineze automat aceste estimări, folosind un proces de învățare din exemple reale de vorbire.

Ceea ce face modulul de recunoaştere a modelelor este să computeze probabilitatea (respectiv posibilitatea bazată pe întreaga experiență anterioară) ca modelul răspunzător de recunoaştere să fie, de fapt, reprezentat pe moment de imput-urile sale active. Fiecare input ajuns la modul devine activ dacă recognizatorul corespunzător de nivel inferior emite un impuls (ceea ce înseamnă că modelul de nivel inferior a fost recunoscut). Fiecare input codifică, de asemenea, mărimea observată (pentru o dimensiune adecvată, cum ar fi întinderea temporală, magnitudinea fizică sau un alt continuum), astfel încât mărimea să poată fi comparată de modul, la computarea probabilității globale a modelului, cu parametrii de mărime stocați pentru fiecare input.

Cum poate creierul (şi cum poate un sistem de inteligență artificială) să computeze probabilitatea globală (pe care modulul este răspunzător să o recunoască) ca modelul să fie prezent, date fiind (1) input-urile (fiecare cu o mărime observată), (2) parametrii stocați de mărime (mărimea aşteptată şi

variabilitatea mărimii) pentru fiecare input şi (3) parametrii de importanţă ai fiecărui input? În anii 1980 şi 1990, împreună cu alţii, am fost pionierul unei metode matematice denumite Modelele Markov Ierarhice Ascunse pentru învăţarea acestor parametri şi pentru folosirea lor ulterioară în vederea recunoaşterii modelelor ierarhice. Am folosit această tehnică în recunoaşterea vorbirii umane şi în înţelegerea limbajului uman. Descriu această abordare mai detaliat în Capitolul 7.

Revenind la fluxul de recunoaştere de la un anumit nivel de recognizatori de modele, în exemplul anterior vedem informaţia curgând în sus pe ierarhia conceptuală, de la caracteristicile elementare ale literelor la cuvinte. Recunoaşterea va continua să se propage de aici mai sus, la fraze şi apoi la structuri şi mai complexe de limbaj. Dacă urcăm alte câteva duzini de niveluri, ajungem la concepte de nivel superior, cum ar fi ironia şi invidia. Deşi toţi recognizatorii de modele funcţionează simultan, este necesar un anumit timp pentru ca recunoaşterea să avanseze în această ierarhie conceptuală. Traversarea fiecărui nivel are nevoie pentru procesare de câteva sutimi până la câţevă zecimi de secundă. Experimentele au dezvăluit că un model de nivel moderat înalt, cum ar fi un chip, are nevoie pentru recunoaştere de cel puţin o zecime de secundă. Este posibil ca procesul să dureze până la o secundă întreagă în cazul în care intervin distorsiuni semnificative. Dacă ar fi secvenţial (cum sunt computerele convenţionale) şi ar realiza recunoaşterea secvenţializată a fiecărui model, creierul ar trebui să ia în calcul toate modelele posibile de nivel inferior înainte să treacă la următorul nivel. Prin urmare, ar fi necesare milioane de cicli doar pentru a se trece prin fiecare nivel. Exact asta se întâmplă atunci când simulăm asemenea procese pe un computer. În orice caz, reţine, cititorule, că un computer procesează de milioane de ori mai rapid decât circuitele noastre biologice.

Un aspect foarte important care trebuie menţionat aici este acela că informaţia se deplasează şi în jos în ierarhia conceptuală, nu numai în sus. Această deplasare în jos este şi mai semnificativă. Dacă, de exemplu, citim de la stânga la dreapta şi am văzut şi am recunoscut deja literele „A," „P", „P" şi „L", recognizatorul pentru „APPLE" anticipează că este posibil să vadă un „E" în următoarea poziţie. Recognizatorul trimite un semnal în jos, la recognizatorul pentru „E", care spune, de fapt: „Te rog să fii atent, că e o mare probabilitate să vezi modelul tău «E» cât de curând, aşa că stai cu ochii pe el". Recognizatorul pentru „E" îşi ajustează atunci pragul astfel încât să fie mult mai posibil să recunoască un „E". Prin urmare, dacă în continuare apare

o imagine care e vag asemănătoare cu un „E", dar poate că este ştearsă, astfel încât n-ar fi fost recunoscută ca „E" în condiții „normale", recognizatorul de „E" poate, totuşi, să indice că a văzut într-adevăr un „E", pentru că se aştepta la asta.

Prin urmare, neocortexul anticipează ce se aşteaptă să întâlnească. Anticiparea viitorului este unul dintre principalele motive pentru care avem neocortex. La nivelul conceptual cel mai înalt, facem în permanență predicții – cine va intra primul pe uşă, ce e posibil să spună o persoană în continuare, ce ne aşteptăm să vedem după ce trecem de un colț, rezultatele posibile ale acțiunilor noastre şi aşa mai departe. Aceste predicții survin în permanență, la fiecare nivel al ierarhiei neocortexului. Adeseori identificăm greşit oameni, obiecte şi cuvinte, pentru că pragul nostru de confirmare a unui model aşteptat este prea redus.

Semnalelor pozitive li se adaugă semnale negative sau inhibitoare, care indică faptul că e mai puțin probabil să existe un anumit model. Aceste semnale vin de la nivelurile conceptuale mai joase (de exemplu, recunoaşterea unei mustăți va inhiba probabilitatea ca persoana pe care o observ că stă la un rând să fie soția mea), dar şi de la niveluri superioare (de exemplu, ştiu că soția mea este plecată într-o călătorie, deci persoana de la rând nu poate fi ea). Când un recognizator de modele primeşte un semnal inhibitor, ridică pragul de recunoaştere, dar este în continuare posibil ca modelele să emită impulsuri (deci, dacă persoana de la rând e într-adevăr soția mea, aş putea s-o recunosc).

Natura datelor care vin la recognizatorul neocortical de modele

În continuare, să luăm în considerare felul cum arată datele pentru un model. Dacă modelul este un chip, datele există în cel puțin două dimensiuni. Nu putem spune că ochii sunt neapărat primii, că sunt urmați de nas şi aşa mai departe. Aceeaşi observație este valabilă şi în cazul majorității sunetelor. O piesă muzicală are cel puțin două dimensiuni. Pot să existe mai multe instrumente şi/sau voci care emit sunete în acelaşi timp. Mai mult, o singură notă a unui instrument complex cum ar fi pianul este compusă din mai multe frecvențe. O singură voce umană posedă, simultan, niveluri variabile de energie în zeci de benzi de frecvență diferite. Prin urmare, un model

de sunet poate fi complex în orice moment şi aceste momente complexe se întind în timp. Input-ul tactil este de asemenea bidimensional, deoarece pielea e un organ de simţ bidimensional şi asemenea modele se pot schimba în funcţie de cea de-a treia dimensiune, timpul.

Aşadar, s-ar părea că input-ul pentru un procesor neocortical de modele trebuie să conţină modele bi- dacă nu chiar tridimensionale. Pe de altă parte, putem observa în structura neocortexului că input-urile de model sunt doar liste unidimensionale. Toate cercetările noastre în domeniul creării sistemelor artificiale de recunoaştere a modelelor (cum ar fi sistemele de recunoaştere a vorbirii şi cele de recunoaştere vizuală) demonstrează că putem (şi am făcut acest lucru) să reprezentăm fenomene bi- şi tridimensionale cu asemenea liste unidimensionale. Voi descrie felul cum funcţionează aceste metode în Capitolul 7, dar deocamdată putem continua cu înţelegerea faptului că input-ul pentru fiecare procesor de modele este o listă unidimensională, cu toate că modelul însuşi poate să reflecte în mod inerent mai multe dimensiuni.

Ar trebui să luăm în calcul în acest moment ideea intuitivă că modelele pe care le-am învăţat pentru recunoaştere (de exemplu, un anumit câine sau ideea generală de „câine", o notă muzicală sau o piesă muzicală) ilustrează exact acelaşi mecanism care stă la baza amintirilor noastre. Amintirile noastre sunt de fapt modele organizate ca liste (în care fiecare componentă, de pe fiecare listă, este un alt model din ierarhia corticală), pe care le-am învăţat şi apoi le recunoaştem atunci când venim în contact cu stimulul adecvat. De fapt, amintirile există în neocortex tocmai pentru a fi recunoscute.

Singura excepţie se află la nivelul conceptual cel mai de jos, la care datele de input pentru un model reprezintă informaţii senzoriale specifice (de exemplu, date în imagini de la nervul optic). Chiar şi acest cel mai de jos nivel al modelului, pe de altă parte, s-a transformat în mod semnificativ în modele simple înainte să ajungă la cortex. Lista de modele care constituie o amintire are o ordine strictă, aşa că suntem capabili să ne reamintim aceste modele numai în respectiva ordine, de unde şi dificultăţile pe care le avem în inversarea amintirilor.

Amintirea trebuie să fie declanşată de un alt gând/amintire (ele reprezintă unul şi acelaşi lucru). Putem trăi experienţa mecanismului de declanşare în clipa în care percepem un model. Când am perceput „A", „P", „P" şi „L", modelul „A P P L E" – „măr" – a anticipat că urma să vedem un „E"

şi a declanşat modelul „E" care era aşteptat. Prin urmare, cortexul nostru „gândeşte" că vede un „E" chiar înainte să-l vadă. Dacă această interacţiune din cortexul nostru ne reţine atenţia, ne vom gândi la un „E" chiar înainte să-l vedem sau chiar dacă nu-l vedem deloc. Un mecanism similar ne declanşează amintiri vechi. De obicei este vorba despre un întreg lanţ de asemenea legături. Chiar dacă posedăm un anumit nivel de conştientizare a amintirilor (respectiv modelele) care au declanşat amintirea veche, amintirile (modelele) nu au etichete în limbaj sau imagini. Aceasta este conştientizarea. Acesta este motivul pentru care amintirile vechi pot părea că ne răsar brusc în conştiinţă. Fiind îngropate şi neactivate vreme de ani întregi probabil, au nevoie de un declanşator, tot aşa cum o pagină web are nevoie de un link web pentru a fi activată. Şi, exact aşa cum o pagină web poate ajunge „orfană" din cauză că nicio altă pagină nu conţine un link pentru ea, tot aşa se poate întâmpla şi cu amintirile noastre.

Gândurile ne sunt activate în mare măsură în unul sau două moduri, nedirecţionat şi direcţionat, ambele folosind aceleaşi link-uri corticale. În modul nedirecţionat, lăsăm link-urile să se joace, fără să încercăm să le orientăm spre vreo direcţie anume. Unele forme de meditaţie (cum ar fi meditaţia transcendentală, pe care o practic) se bazează pe lăsarea minţii să facă exact acest lucru. Visele au şi ele aceeaşi proprietate.

În cazul gândirii direcţionate, încercăm să străbatem un proces mult mai ordonat de reapelare a amintirii (o poveste, de exemplu) sau de rezolvare a unei probleme. Iar asta implică parcurgerea listelor din neocortex, însă agitaţia structurată a gândului nedirecţionat va însoţi şi ea procesul. Conţinutul complet al gândirii noastre este, prin urmare, foarte dezordonat, fenomen pe care James Joyce l-a pus în lumină în romanele sale pe tema „fluxului conştientizării".

Pe măsură ce gândeşti prin amintirile/poveştile/modelele din viaţa ta, indiferent dacă ele implică întâlnirea întâmplătoare cu o mamă care împinge un cărucior cu un bebeluş în el sau povestea mult mai importantă a felului în care ţi-ai cunoscut soţia/soţul, amintirile rezidă într-o succesiune de modele. Deoarece aceste modele nu sunt etichetate cu cuvinte, sunete, imagini sau video, când încerci să-ţi aminteşti un eveniment semnificativ, în esenţă reconstruieşti în minte modelele, deoarece imaginile propriu-zise nu există.

Dacă ai vrea să „citeşti" gândurile unei persoane şi să arunci o privire asupra a ceea ce se petrece mai exact în neocortexul persoanei respective,

ar fi foarte dificil să-i interpretezi amintirile, chiar dacă ai încerca să vizualizezi modelele stocate în neocortex în aşteptarea momentului când vor fi declanşate sau pe cele care au fost declanşate deja şi a căror experienţă este trăită în momentul respectiv sub formă de gânduri active. Ceea ce am „vedea" ar fi activarea simultană a milioane de recognizatori de modele. O sutime de secundă mai târziu, am vedea un set diferit, cu un număr comparabil de recognizatori de modele activaţi. Fiecare asemenea model ar reprezenta o listă a celorlalte modele şi fiecare dintre aceste modele ar fi o listă a celorlalte modele şi aşa mai departe, până am ajunge la modelele elementare, de la nivelul cel mai de jos. Ar fi extrem de dificil să interpretăm semnificaţia modelelor de nivel superior fără *să copiem* în propriul nostru cortex toată informaţia de la fiecare nivel. În acest fel, fiece model din neocortexul nostru este plin de semnificaţii numai în lumina întregii informaţii deplasate la nivelurile de sub el. Mai mult de-atât, alte modele de la acelaşi nivel şi de la niveluri superioare sunt de asemenea relevante pentru interpretarea unui anumit model, deoarece asigură contextul. Prin urmare, adevărata „citire a gândurilor" ar necesita nu doar detectarea activărilor axonilor relevanţi din creierul unei persoane, ci şi examinarea întregului neocortex cu toate amintirile sale, pentru înţelegerea acestor activări.

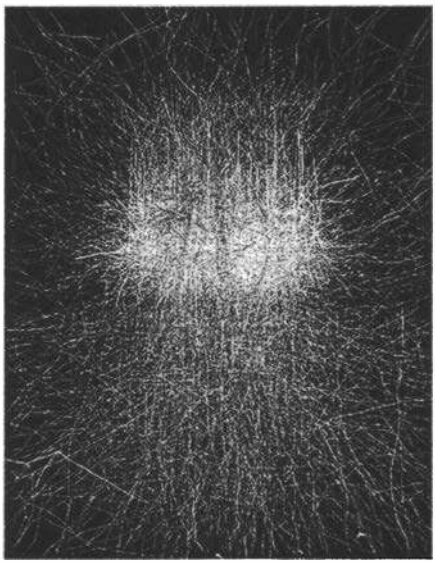
Când trăim experienţa propriilor gânduri şi amintiri, „ştim" ce înseamnă acestea, dar ele nu există ca gânduri şi amintiri uşor explicabile. Dacă vrem să le împărtăşim altora, trebuie să le traducem în limbaj. Şi această sarcină este îndeplinită tot de neocortex, el folosind recognizatori de modele antrenate cu modele care au fost învăţate în scopul folosirii limbajului. Limbajul este el însuşi suficient de ierarhic şi evoluat pentru a profita de natura ierarhică a neocortexului, care, la rândul ei, reflectă natura ierarhică a realităţii. Capacitatea înnăscută a omului de a învăţa structurile ierarhice ale limbajului, despre care a scris Noam Chomsky, reflectă structura neocortexului. Într-o lucrare din anul 2002 al cărei coautor a fost, Chomsky citează atributul „revenirii" ca fiind acela care explică facultatea unică a limbajului la specia umană.[4] Revenirea, potrivit lui Chomsky, reprezintă capacitatea de a reuni bucăţi mai mici într-o bucată mare, proces urmat de folosirea bucăţii mari ca parte a unei alte structuri şi continuarea recurentă a acestui proces. În acest fel, suntem capabili să construim structuri elaborate de propoziţii şi paragrafe dintr-un set limitat de cuvinte. Cu toate că Chomsky nu se referă explicit aici la structura creierului, abilitatea pe care o descrie este exact ceea ce face neocortexul.

Speciile inferioare de mamifere își consumă în mare parte neocortexul cu provocările stilurilor lor particulare de viață. Specia umană a căpătat capacități suplimentare dezvoltându-și un cortex substanțial mai mare, cu care poate să se ocupe de limbajul vorbit și scris. Unii oameni au învățat această abilitate mai bine decât alții. Dacă am spune o anumită poveste de multe ori, am începe să învățăm practic succesiunea termenilor care descriu povestea ca pe o serie de succesiuni separate. Chiar și în acest caz, amintirea nu este o succesiune strictă de cuvinte, ci mai curând de structuri de limbaj de care avem nevoie pentru a traduce în succesiuni specifice de cuvinte de fiecare dată când reproducem povestea. Acesta este motivul pentru care spunem povestea puțin diferit de fiecare dată când o împărtășim (dacă nu învățăm succesiunea exactă a cuvintelor ca model).

Pentru fiecare dintre aceste descrieri ale proceselor specifice de gândire, mai trebuie să luăm în considerare și subiectul redundanței. După cum am menționat, nu avem nici măcar un model care să reprezinte entitățile importante din viața noastră, indiferent dacă aceste entități constituie categorii senzoriale, concepte de limbaj sau amintiri ale unor evenimente. Toate evenimentele importante – la toate nivelurile – se repetă de multe ori. Unele dintre aceste recurențe reprezintă diferite perspective și puncte de avantaj. Acesta este unul dintre principalele motive pentru care putem recunoaște un chip familiar din diverse puncte de observare și în diferite condiții de iluminare. Fiecare nivel din ierarhie are o redundanță substanțială, permițând suficientă variabiliate corespunzătoare acelui concept.

Prin urmare, dacă ar fi să ne imaginăm neocortexul tău când te uiți la o persoană iubită, am vedea o mulțime de impulsuri nervoase generate de axonii recognizatorilor de modele de la toate nivelurile, pornind de la nivelul de bază al modelelor senzoriale primitive și ajungând la numeroasele și diferitele modele ce reprezintă imaginea persoanei iubite. Am sesiza, de asemenea, generarea unui număr foarte mare de impulsuri nervoase reprezentând alte aspecte ale situației, cum ar fi mișcările persoanei respective, cuvintele pe care le spune și așa mai departe. Prin urmare, dacă experiența pare mult mai bogată decât o simplă mișcare sistematică în sus într-o ierarhie de caracteristici, chiar așa și este.

Însă mecanismul de bază al ascensiunii într-o ierarhie de recognizatori de modele în care fiecare model conceptual reprezintă un concept mai abstract și mai integrat rămâne valid. Fluxul de informație în direcție descendentă în ierarhie este și mai mare, fiecare nivel activat al modelului



Simularea computerizată a generării simultane de impulsuri de către mai mulți recognizatori de modele din neocortex

recunoscut trimițând predicții la recognizatorul de modele de la următorul nivel cu privire la ceea ce este posibil să se ivească în continuare. Aparenta bogăție a experienței umane este rezultatul faptului că toate acele sute de milioane de recognizatori de modele din neocortexul nostru își evaluează simultan input-urile.

În Capitolul 5, voi discuta despre fluxul de informație care vine în neocortex de la pipăit, vedere, auz și alte organe senzoriale. Aceste input-uri timpurii sunt procesate de regiuni corticale dedicate tipurilor relevante de input-uri senzoriale (cu toate că există o enormă plasticitate a atribuirii acestor regiuni, reflectând uniformitatea funcției la nivelul neocortexului). Ierarhia conceptuală continuă peste cele mai înalte concepte în fiecare regiune senzorială a neocortexului. Ariile corticale de asociere integrează input-urile diferitelor organe de simț. Când auzim ceva care se aseamănă poate cu vocea soției/soțului și apoi vedem ceva ce poate că-i semnalează prezența, nu ne vom angaja într-un proces elaborat de deducție logică; dimpotrivă, vom percepe instantaneu că e soțul/soția din combinația acestor recogniții senzoriale. Integrăm toate indiciile senzoriale și perceptuale

relevante – poate chiar şi mirosul parfumului sau al coloniei – într-o percepție multinivel.

La un nivel conceptual aflat deasupra ariilor corticale senzoriale de asociere, devenim capabili să ne confruntăm – să percepem, să ne reamintim şi să ne gândim la – cu concepte mult mai abstracte. La nivelul cel mai înalt, recunoaştem modele cum ar fi *E amuzant; E frumuşică; E ironic* şi aşa mai departe. Amintirile noastre includ şi aceste modele de recogniție abstractă. De exemplu, este posibil să ne reamintim că vorbeam cu cineva care a spus ceva amuzant şi am râs, cu toate că s-ar putea să nu ne mai aducem aminte de gluma propriu-zisă. Succesiunea memoriei acestei amintiri a înregistrat pur şi simplu percepția umorului, dar nu şi conținutul exact al comunicării care l-a generat.

În capitolul anterior, am subliniat că ne putem aminti adeseori de un model cu toate că nu-l recunoaştem suficient de bine încât să fim în stare să-l descriem. De exemplu, cred că aş putea să extrag, dintr-un grup de fotografii cu femei, fotografia tinerei cu căruciorul pe care am văzut-o astăzi mai devreme, în ciuda faptului că nu pot s-o vizualizez de fapt şi nu pot s-o descriu prea amănunțit. În acest caz, amintirea mea legată de ea este o listă cu anumite caracteristici de nivel înalt. Aceste caracteristici nu au ataşate etichete de limbaj sau de imagine şi nu sunt imagini pixelate, astfel încât, deşi pot să mă gândesc la femeie, nu pot să o descriu. Pe de altă parte, dacă mi se dă o fotografie a ei, pot procesa imaginea, fapt ce se concretizează în recunoaşterea aceloraşi caracteristici de nivel înalt care au fost identificate prima oară când am văzut-o. Aş putea să determin, prin urmare, că aceste caracteristici corespund şi, astfel, aş alege plin de încredere fotografia ei.

Cu toate că am văzut-o pe această femeie o singură dată, la o plimbare, probabil că există deja còpii multiple ale modelului ei în neocortexul meu. Pe de altă parte, dacă nu mă gândesc la ea o perioadă de timp dată, aceşti recognizatori de modele se vor atenua: volumul de redundanță se va diminua până când anumite amintiri se vor stinge. Totodată, acum, după ce mi-am întipărit-o în minte pe această femeie scriind despre ea, probabil că nu o voi uita aşa de uşor.

Asocierea şi invarianța

În capitolul anterior, am discutat despre modul cum putem recunoaşte un model chiar dacă acesta nu e prezent integral şi, de asemenea, chiar

dacă este distorsionat. Prima capacitate se numeşte autoasociere: capacitatea de asociere a modelului cu o parte din el însuşi. Structura fiecărui recognizator de modele încorporează în mod inerent această capacitate.

Pe măsură ce fiecare input de la un recognizator de model inferior urcă la unul de nivel superior, conexiunea poate să aibă o „greutate" care să indice cât de important este acel element anume din model. Astfel, elementele cele mai semnificative ale unui model cântăresc mai mult atunci când se apreciază că modelul respectiv ar trebui să fie considerat „recunoscut". Este foarte posibil ca barba lui Lincoln, perciunii lui Elvis şi celebrul gest al lui Einstein de a scoate limba să cântărească greu în modelele pe care le ştim, deoarece aparţin acestor personaje reprezentative. Recognizatorul de model computează o probabilitate care ia în calcul parametrii de importanţă. Astfel, probabilitatea de ansamblu este mai redusă dacă lipseşte un element sau mai multe, deşi pragul de recunoaştere poate să fie, totuşi, satisfăcut. După cum am subliniat, computaţia probabilităţii de ansamblu (că este prezent modelul) e mai complicată decât o simplă sumă cântărită în ideea că parametrii de mărime trebuie luaţi şi ei în calcul.

Dacă recognizatorul de model a primit semnal de la un recognizator de nivel superior că modelul lui este „aşteptat", atunci pragul este redus corespunzător (respectiv, e făcut mai uşor de accesat). Reversul: un asemenea semnal poate să se adauge pur şi simplu la totalul input-urilor cântărite, compensând prin aceasta un element-lipsă. Asta se întâmplă la toate nivelurile, astfel încât un model cum ar fi un chip, aflat la mai multe niveluri deasupra celui mai de jos, poate fi recunoscut până şi în condiţiile lipsei mai multor caracteristici.

Capacitatea de recunoaştere a modelelor chiar şi în condiţiile transformării unor aspecte ale acestora se numeşte invarianţa caracteristicilor şi se realizează în patru feluri. În primul rând, există transformări globale care sunt realizate înainte ca neocortexul să primească datele senzoriale. Vom discuta drumul datelor senzoriale de la ochi, urechi şi piele în secţiunea *Calea senzorială*, la pagina 93.

A doua metodă profită de redundanţa memoriei corticale de modele. Mai ales în cazul subiectelor importante, am acumulat multe perspective diferite şi puncte de observare pentru fiecare model. Prin urmare, numeroase variaţii sunt stocate şi procesate separat.

Cea de-a treia şi cea mai eficientă metodă ţine de abilitatea combinării a două liste. Una dintre liste poate să conţină un set de transformări despre

care știm că se pot aplica unei anumite categorii de modele; cortexul aplică exact această listă de posibile schimbări la un alt model. În acest fel înțelegem fenomene cum ar fi metaforele și comparațiile.

De exemplu, am învățat că anumite foneme (sunetele de bază ale unei limbi) pot să lipsească din limba vorbită (de exemplu, „goin'"). Dacă învățăm pe urmă un alt cuvânt vorbit (de exemplu „driving"), probabil că vom putea să recunoaștem acest cuvânt și când lipsește unul dintre fonemele sale, în ciuda faptului că nu am mai avut experiența respectivului cuvânt în această formă, deoarece ne-am familiarizat cu fenomenul general al omiterii anumitor foneme. Alt exemplu: putem învăța că unui anumit artist îi place să scoată în evidență (făcându-le mai mari) anumite elemente ale feței, cum ar fi nasul. Putem atunci identifica un chip cu care suntem familiarizați și căruia i-a fost aplicată această modificare chiar dacă nu am mai văzut modificarea respectivă aplicată acelui chip.

Anumite modificări artistice scot în evidență chiar caracteristicile care sunt recunoscute de neocortexul nostru bazat pe recunoașterea modelelor. După cum am menționat, exact acesta este fundamentul caricaturii.

A patra metodă rezultă din parametrii de mărime care-i permit unui singur modul să codifice mai multe aspecte ale unui model. De exemplu, am auzit de multe ori cuvântul „steep" – „abrupt". Un anumit modul de recunoaștere a modelului care identifică acest cuvânt rostit poate codifica multiple exemple indicând că durata lui [E] are o înaltă variabilitate de așteptări. Dacă toate modulele alocate cuvintelor care conțin [E] împărtășesc un fenomen similar, această variabilitate poate fi codificată chiar în modelele lui [E]. Pe de altă parte, diferitele cuvinte care conțin [E] (sau multe alte foneme) pot să aibă variabilități așteptate diferite. De exemplu, cuvântul „peak" – „vârf" – e posibil să nu conțină fonemul [E] la fel de evident cum o face cuvântul „steep".

Învățarea

> Oare nu ne creăm noi înșine succesorii la supremația pe Pământ, contribuind zilnic la frumusețea și delicatețea organizării lor, dându-le zilnic mai multă îndemânare și punându-le la dispoziție din ce în ce mai mult din acea forță automată și autoreglată care va fi mai bună decât orice intelect?
>
> **— Samuel Butler, 1871**

Principalele activități ale creierului constau în a se modifica pe el însuși.
— **Marvin Minsky, Societatea Minții**

Până acum am examinat modul în care recunoaștem (senzorial și perceptiv) modele și ne amintim succesiuni de modele (amintirile legate de lucruri, oameni și evenimente). Pe de altă parte, nu ne-am născut cu un neocortex plin cu toate aceste modele. Neocortexul nostru este un teritoriu virgin în momentul când ia naștere creierul. Are capacitatea de a învăța și prin urmare de a crea conexiuni între recognizatorii de modele, însă dobândește aceste conexiuni prin experiență.

Procesul de învățare începe chiar înainte de a ne naște, producându-se concomitent cu procesul biologic de creștere propriu-zisă a creierului. Fătul are deja creier la vârsta de o lună, cu toate că, în esență, acesta este un creier reptilian, fătul parcurgând de fapt, în uter, la mare viteză, evoluția biologică. Creierul devine în mod distinct un creier uman cu un neocortex uman înainte să ajungă în al treilea trimestru al sarcinii.

În acest moment, fătul trăiește experiențe, iar neocortexul învață. Aude sunete, mai ales bătăile de inimă ale mamei, acesta fiind, foarte posibil, un motiv pentru care calitățile ritmice ale muzicii sunt universale în cultura omului. Absolut toate civilizațiile descoperite vreodată au avut ca parte a culturii lor muzica, spre deosebire de alte forme ale artei, cum ar fi arta plastică. De asemenea, este și motivul pentru care ritmul muzicii este comparabil cu ritmul bătăilor inimii. În mod evident, ritmurile muzicii variază – altminteri, muzica nu ne-ar mai reține interesul –, dar bătăile inimii variază și ele. Cele excesiv de regulate reprezintă, de fapt, un simptom de boală cardiacă. Ochii fătului sunt parțial deschiși la douăzeci și șase de săptămâni de la concepere și sunt deschiși complet aproape tot timpul înainte de a douăzeci și opta săptămână de la concepere. Poate că nu sunt prea multe de văzut în uter, dar există modele de întuneric și lumină pe care neocortexul începe să le proceseze.

Așadar, deși nou-născutul beneficiază în uter de o mică experiență, aceasta este clar limitată. De asemenea, neocortexul poate să învețe de la creierul vechi (un subiect pe care îl discut în Capitolul 5), dar, în general, la naștere, copilul are multe de învățat – aproape totul, de la sunetele și formele primitive de bază, până la metafore și sarcasm.

Învățarea este fundamentală pentru inteligența umană. Dacă ar trebui să realizăm un model perfect ca să simulăm neocortexul omului (așa cum

încearcă să facă Proiectul Blue Brain) şi toate celelalte regiuni de care creierul are nevoie pentru a funcționa (de pildă hipocampul şi talamusul), acest neocortex artificial nu ar putea face prea multe – la fel cum nou-născutul nu poate face prea multe (în afară de a fi drăgălaş, ceea ce reprezintă în mod cert o adaptare-cheie pentru supraviețuire).

Învățarea şi recunoaşterea au loc simultan. Începem să învățăm imediat şi, de îndată ce am învățat un model, începem să-l recunoaştem. Neocortexul încearcă permament şi continuu să găsească logica input-ului care îi este pus la dispoziție. Dacă un anumit nivel se dovedeşte incapabil să proceseze complet şi să recunoască un model, acesta este trimis la următorul nivel superior. Dacă niciunul dintre niveluri nu reuşeşte să recunoască modelul, acesta este considerat nou. Clasificarea unui model ca fiind nou nu înseamnă în mod necesar că toate aspectele sale sunt noi. Dacă ne uităm la picturile unui anumit artist şi vedem o pisică având o trompă de elefant, putem identifica fiecare dintre trăsăturile caracteristice, dar vom observa că acest model combinat este ceva nou şi că e posibil să ni-l reamintim. Nivelurile conceptuale superioare ale neocortexului, care înțeleg contextul – de exemplu, circumstanța că această pictură este un exemplu al operei unui anumit artist şi că participăm la deschiderea unei expoziții de tablouri noi ale respectivului artist –, vor constata combinația neobişnuită de modele din chipul pisicii-elefant, dar în acelaşi timp vor include aceste detalii contextuale ca modele suplimentare de amintire.

Amintirile noi, cum ar fi chipul pisicii-elefant, sunt stocate într-un recognizator de modele disponibil. Hipocampul joacă un rol în acest proces şi vom discuta ce se ştie despre mecanismele biologice propriu-zise în capitolul următor. Pentru scopurile modelului nostru de neocortex, este suficient să spun că modelele care nu sunt recunoscute sunt stocate ca modele noi şi sunt conectate în mod adecvat la modelele de nivel inferior care le formează. Chipul pisicii-elefant, de exemplu, va fi stocat în mai multe moduri. Noul aranjament de componente ale feței va fi stocat ca o amintire contextuală care îl include şi pe artist, şi situația ori poate faptul că am râs când l-am văzut prima oară.

Amintirile care sunt recunoscute cu succes pot să se concretizeze şi în crearea unui nou model de atingere a unei redundanțe mai mari. Dacă modelele nu sunt recunoscute perfect, este posibil să fie stocate ca şi cum ar reflecta o perspectivă diferită a subiectului care a fost recunoscut.

Şi atunci, care e metoda general valabilă pentru a stabili ce modele sunt stocate? În termeni matematici, problema poate fi formulată după

cum urmează: Folosind limitele disponibile de stocare a modelelor, cum putem reprezenta optim modelele de input cu care ne-am confruntat până acum? Cu toate că este logic să se accepte un anumit nivel de redundanță, nu ar fi practic să fie umplută întreaga arie de stocare (respectiv întregul neocortex) cu modele repetate, de vreme ce asta nu ar permite o suficientă diversitate a modelelor. Un model cum ar fi cel al fonemului [E] în cuvintele rostite este unul de care ne-am lovit de nenumărate ori. E un model simplu de frecvențe sonore şi fără îndoială că se bucură de o redundanță semnificativă în neocortexul nostru. Am putea să ne umplem întregul neocortex cu modele repetate ale fonemului [E]. Există, totuşi, o limită pentru redundanța utilă şi un model obişnuit cum este acesta în mod clar a atins-o.

Există o soluție matematică la această problemă de optimizare: programarea liniară. Aceasta rezolvă alocarea cea mai bună cu putință a resurselor limitate (în cazul de față, un număr limitat de recognizatori de modele) care ar reprezenta toate cazurile cu care s-a pregătit sistemul. Programarea liniară este concepută pentru sisteme cu input-uri unidimensionale, acesta fiind un alt motiv pentru care e optim să reprezentăm input-ul fiecărui modul de recunoaştere de model ca pe un şir liniar de input-uri. Putem folosi această abordare matematică în cazul unui sistem software şi, cu toate că creierul real este limitat şi mai mult de conexiunile fizice pe care le are la dispoziție şi pe care le poate adapta între recognizatorii de modele, metoda este totuşi similară.

O importantă implicație a acestei soluții optime este că experiențele rutiniere sunt recunoscute, dar nu se concretizează în realizarea unei amintiri definitive. În privința plimbării mele, m-am confruntat cu milioane de modele la fiecare nivel, de la muchiile şi nuanțele vizuale de bază, la obiecte cum ar fi stâlpii de iluminat, cutiile poştale, oamenii, animalele şi plantele pe lângă care am trecut pe drum. Aproape nimic de pe parcursul acestei experiențe nu a fost ceva unic şi modelele pe care le-am recunoscut îşi atinseseră de mult nivelul optim de redundanță. Rezultatul este că nu-mi amintesc aproape nimic din cursul acestei plimbări. Cele câteva detalii de care îmi aduc aminte e posibil să fie înlocuite cu noi modele înainte să fac câteva zeci de noi plimbări, exceptând faptul că am memorat această plimbare anume scriind despre ea.

Un aspect important care se aplică atât neocortexului nostru biologic, cât şi tentativelor de a-l reproduce este acela că e dificil să înveți simultan

mai multe niveluri conceptuale. În esență, putem învăța unul, maximum două asemenea niveluri în același timp. Odată ce învățarea devine relativ stabilă, putem trece la asimilarea următorului nivel. Putem continua să operăm reglajul fin al învățării pe nivelurile inferioare, dar învățarea se concentrează pe următorul nivel de abstractizare. Acest lucru este adevărat atât la începutul vieții, când nou-născutul se luptă cu formele de bază, cât și mai târziu în timpul vieții, când ne luptăm să învățăm o altă materie, pe rând, pe câte un nivel de complexitate odată. Descoperim același fenomen la mașinile care imită neocortexul. Pe de altă parte, dacă suntem confruntați cu un material din ce în ce mai abstract, la un moment dat mașina este capabilă să învețe la fel ca omul (deși deocamdată nu pe tot atât de multe niveluri conceptuale).

Output-ul unui model poate face legătura cu un model de la un nivel inferior sau chiar cu modelul propriu-zis, conferindu-i creierului uman puternica abilitate recursivă. Elementul unui model poate fi un punct de decizie bazat pe un alt model. Acest fapt este deosebit de util pentru listele care compun acțiuni – de exemplu, cumpărarea unui nou tub de pastă de dinți dacă actualul este gol. Aceste condiționări există la toate nivelurile. După cum știe oricine a încercat să programeze o procedură, condiționalele sunt vitale pentru descrierea cursului unei acțiuni.

Limbajul gândirii

Visul funcționează ca o supapă de siguranță pentru creierul supraîncărcat.
— Sigmund Freud, *Interpretarea viselor,* 1911

Creierul: un aparat cu care gândim că gândim.
— Ambrose Bierce, *Dicționarul diavolului*

Ca să rezum ceea ce am aflat până acum despre modul de funcționare a neocortexului, te rog să te uiți peste diagrama modulului neocortical de recunoaștere a modelelor de la pagina 45.

a. Dendritele pătrund în modulul care reprezintă modelul. Deși modelele pot părea că au o natură bi- sau tridimensională, sunt reprezentate de o succesiune de semnale unidimensionale. Modelul trebuie

să fie prezent în această ordine (secvențială) pentru ca recognizatorul de modele să fie capabil să-l recunoască. Fiecare dendrită este conectată în primul rând la unul sau la mai mulți axoni ai recognizatorilor de modele, la nivelul conceptual cel mai redus care a recunoscut un model de nivel inferior ce constituie o parte din acest model. Pentru fiecare dintre modelele din input, pot exista mai mulți recognizatori de modele de nivel inferior, care pot genera semnalul că modelul de nivel inferior a fost recunoscut. Pragul necesar pentru recunoaşterea modelului poate fi asigurat chiar dacă nu toate input-urile au transmis semnale. Modulul computează probabilitatea de a fi prezent modelul de care este răspunzător. Această computație ia în calcul parametrii „importanță" şi „mărime" (vezi [f] în continuare).

Reține că unele dendrite transmit semnale spre modul, iar altele dinspre modul. Dacă toate dendritele de input ale acestui recognizator de modele semnalează că modelele lor de nivel redus au fost recunoscute, cu excepția a unul sau două, recognizatorul va transmite un semnal în jos la recognizatorul(-oarele) de modele, recunoscând modelele de nivel inferior care nu au fost încă recunoscute şi indicând faptul că există o mare probabilitate ca acele modele să fie recunoscute curând, aşa că recognizatorul(-rii) de nivel redus ar trebuie să îl aştepte.

b. Când acest recognizator de modele îşi recunoaşte modelul (pe baza activării tuturor sau a majorității semnalelor de input de la dendrite), se activează axonul (output) acestui recognizator de modele. La rândul lui, acest axon poate să se conecteze la o întreagă rețea de dendrite conectate la recognizatori de modele de nivel superior, la care este transmis acest model ca input. Acest semnal transmite magnitudinea informației astfel încât recognizatorii de modele de la următorul nivel superior conceptual să-l poată lua în considerare.

c. Dacă un recognizator de modele de nivel superior primeşte un semnal pozitiv de la toate sau de la majoritatea modelelor constituente cu excepția celui reprezentat de un alt recognizator de modele, recognizatorul de nivel superior poate să transmită un semnal la respectivul recognizator, prin care să-l anunțe că modelul său este aşteptat. Un asemenea semnal ar determina recognizatorul de model să îşi reducă pragul, ceea ce înseamnă că ar fi mai posibil să transmită la axonul lui un semnal care să indice că modelul este recunoscut chiar dacă o parte dintre input-urile sale lipsesc sau sunt neclare.

d. Semnalele inhibitoare de dedesubt ar reduce posibilitatea ca acest recognizator de modele să-şi recunoască modelul. Asta poate rezulta din identificarea modelelor de nivel inferior care sunt inconsistente cu modelul asociat acestui recognizator de modele (de exemplu, recunoaşterea unei mustăți de către un recognizator de nivel inferior ar face mai puțin probabil ca imaginea respectivă să fie „soția mea").

e. Semnalele inhibitorii de deasupra ar face de asemenea mai puțin probabil faptul ca acest recognizator de modele să îşi recunoască modelul. Această situație poate să rezulte dintr-un context de nivel superior care este inconsecvent cu modelul asociat acestui recognizator.

f. Pentru fiecare input, sunt stocați parametri legați de importanță, de mărimea aşteptată şi de variabilitatea aşteptată a mărimii. Pe baza tuturor acestor parametri şi a semnalelor curente care indică input-urile prezente şi magnitudinile lor, modulul computează o probabilitate de ansamblu ca modelul să fie prezent. Un mijloc matematic optim de asigurare a acestui rezultat îl reprezintă *Modelele Markov Ascunse*. Când asemenea modele sunt organizate într-o ierarhie (aşa cum se întâmplă în neocortex sau în tentativele de simulare a neocortexului), le numim Modele Markov Ascunse Ierarhic.

Modelele generate de neocortex generează alte modele. Modelele parțial complete transmit semnale în jos în ierarhia conceptuală; modelele complete transmit semnale în sus în respectiva ierarhie. Aceste modele neocorticale constituie limbajul gândurilor. La fel ca limbajul, şi ele sunt ierarhizate, dar nu sunt un limbaj ca atare. Gândurile noastre nu sunt concepute în principal în elemente ale limbajului, cu toate că, de vreme ce limbajul există în neocortexul nostru şi ca ierarhii de modele, putem avea gânduri bazate pe limbaj. Însă, în cea mai mare parte a lor, gândurile sunt reprezentate în aceste modele neocorticale.

Aşa cum am discutat mai înainte, chiar dacă reuşeam să detectăm activările de model din neocortexul cuiva, tot am fi avut prea puțină idee despre ce înseamnă acele activări de model în lipsa accesului la întreaga ierarhie de modele de deasupra şi de dedesubtul fiecărui model activat. Acest fapt ar presupune, într-o măsură destul de mare, accesul la întregul neocortex al persoanei respective. Ne este dificil să înțelegem conținutul propriilor gânduri, iar înțelegerea gândurilor altcuiva necesită stăpânirea unui alt neocortex, străin. Desigur, încă nu avem acces la neocortexul altei

persoane; trebuie, prin urmare, să ne bazăm pe încercările acelei persoane de a-şi exprima gândurile prin limbaj (precum şi prin alte mijloace, cum ar fi gesturile). Abilitatea incompletă a omului de a-şi îndeplini aceste necesităţi de comunicare mai adaugă un strat de complexitate – nu e de mirare că ne înţelegem unii pe alţii atât de greşit.

Avem două moduri de gândire. Unul îl reprezintă gândirea indirectă, în cadrul căreia gândurile se generează unele pe altele în mod nonlogic. Când ne confruntăm cu rememorarea bruscă a unei amintiri din urmă cu ani sau decenii în timp ce facem altceva, de pildă greblăm frunze sau mergem pe stradă, experienţa este rememorată – aidoma tuturor amintirilor – ca o succesiune de modele. Nu vizualizăm imediat scena decât dacă putem reapela o sumedenie de amintiri care ne permit să sintetizăm o amintire mai robustă. Dacă izbutim să vizualizăm o scenă în acest mod, în esenţă o recreăm în mintea noastră din indicii, în momentul reamintirii; amintirea propriu-zisă nu este stocată sub formă de imagini sau de vizualizări. După cum am menţionat anterior, generatorii care determină acest gând să ne răsară în minte pot să fie sau pot să nu fie evidenţi. E posibil ca succesiunea gândurilor relevante să fi fost dată uitării pe loc. Chiar dacă ne reamintim o scenă, aceasta va reprezenta o succesiune neliniară şi ocolită de asocieri.

Al doilea mod de a gândi este gândirea direcţionată, la care apelăm atunci când încercăm să rezolvăm o problemă sau să formulăm un răspuns organizat. De exemplu, este posibil să repetăm în gând ceea ce intenţionăm să-i spunem cuiva sau să formulăm un pasaj pe care vrem să-l scriem (chiar şi într-o carte mintală). Când ne gândim la sarcini de acest gen, deja le-am descompus pe rând într-o ierarhie de subsarcini. Scrierea unei cărţi, de exemplu, presupune scrierea unor capitole; fiecare capitol conţine secţiuni; fiecare secţiune conţine paragrafe; fiecare paragraf conţine propoziţii care exprimă idei; fiecare idee are propria configurare de elemente; fiecare element şi fiecare relaţie dintre elemente este o idee şi ea trebuie articulată – şi aşa mai departe. Totodată, structurile noastre neocorticale au învăţat anumite reguli care ar trebui respectate. Dacă sarcina este scrisul, atunci s-ar impune să evităm repetiţia inutilă; ar trebui să încercăm să ne asigurăm că cititorul poate urmări ce citeşte; ar trebui să ne străduim să respectăm reguli de gramatică şi stil şi aşa mai departe. Scriitorul trebuie, prin urmare, să-şi construiască în minte un model al cititorului, iar o asemenea construcţie este de asemenea ierarhică. Prin gândirea direcţionată, străbatem listele în neocortex, fiecare dintre ele extinzându-se în ierarhii vaste de subliste,

fiecare cu propriile considerente. Nu uita că elementele de pe o listă aflată într-un model neocortical pot include condiționări, așa încât gândurile și acțiunile noastre ulterioare vor depinde de evaluările făcute pe măsură ce străbatem procesul.

Mai mult de-atât, fiecare dintre gândurile direcționate va genera ierarhii de gânduri nedirecționate. Un continuu iureș de reflectare ne slujește atât experiențele senzoriale, cât și încercările de gândire direcționată. Experiența noastră mentală propriu-zisă este complexă și amestecată, fiind compusă din furtuni cu fulgere de modele generate, care se schimbă de aproximativ o sută de ori pe secundă.

Limbajul viselor

Visele sunt exemple de gânduri nedirecționate. Au o anumită logică, deoarece fenomenul de generare a unui gând de către alt gând se bazează pe conexiuni reale între modele la nivelul neocortexului nostru. În măsura în care visul nu este logic, încercăm să-l reparăm prin capacitatea noastră de a născoci. Așa cum voi demonstra în Capitolul 9, pacienții cu creierul secționat (al căror corp calos, care conectează cele două emisfere ale creierului, este secționat sau lezat) născocesc (inventează) explicații cu ajutorul creierului stâng – care controlează centrul vorbirii – pentru a lămuri ceea ce tocmai a făcut creierul drept cu un input la care creierul stâng nu a avut acces. Născocim în permanență, explicând rezultatul evenimentelor. Dacă vrei un bun exemplu în acest sens, caută comentariile cotidiene pe tema mișcărilor de pe piețele financiare. Indiferent cum performează piețele, este întotdeauna posibil să se născocească o explicație bună pentru motivul care a dus la efectul respectiv; iar asemenea comentarii post factum există din abundență. Desigur, dacă numeroșii comentatori ar înțelege cu adevărat piețele, nu ar trebui să-și irosească timpul făcând comentarii.

Actul născocirii este realizat, desigur, și de neocortex, care se pricepe să inventeze povești și explicații capabile să satisfacă anumite constrângeri. Facem acest lucru de fiecare dată când spunem din nou o poveste. Includem detalii care poate că nu ne sunt accesibile sau pe care poate că le-am uitat, în așa fel încât povestea să fie mai logică. Acesta este motivul pentru care poveștile se schimbă cu timpul pe măsură ce sunt spuse de nenumărate ori de noi povestitori, care poate că au interese diferite. Pe de altă parte, la

fel cum limbajul vorbit a dus la limbajul scris, tot aşa am avut la dispoziţie o tehnologie care a putut înregistra versiunea definitivă a unei poveşti şi a putut împiedica acest gen de derivă.

Conţinutul propriu-zis al unui vis, în măsura în care ni-l mai reamintim, este iarăşi o succesiune de modele. Aceste modele reprezintă constrângerile dintr-o povestire; pe urmă născocim o poveste care să corespundă acestor constrângeri. Versiunea visului pe care o povestim (fie şi nouă înşine, muteşte) este chiar această născocire. Când povestim un vis, declanşăm cascade de modele care umplu visul real, aşa cum l-am trăit de fapt.

Între gândurile din vise şi gândirea din starea de veghe există o deosebire-cheie. Una dintre lecţiile pe care le-am învăţat în viaţă e că anumite acţiuni, chiar şi gânduri, nu sunt permise în lumea reală. De exemplu, am învăţat că nu putem să ne îndeplinim imediat visele. Există reguli împotriva smulgerii banilor-gheaţă din sertarul casierului de la magazin şi constrângeri în interacţiunile cu o persoană de care putem fi atraşi fizic. De asemenea, am învăţat că anumite gânduri nu sunt permise pentru că sunt interzise de cultură. Pe măsură ce ne însuşim deprinderi profesionale, învăţăm şi modurile de gândire recunoscute şi recompensate în profesiile noastre şi prin asta evităm modelele de gândire care ar putea să contravină metodelor şi normelor respectivei profesii. Multe dintre aceste tabùuri sunt meritorii, deoarece asigură ordinea socială şi consolidează progresul. Pe de altă parte însă, ele pot şi să împiedice progresul, prin impunerea unei ortodoxii neproductive. Einstein a vrut să lase în urmă exact această ortodoxie atunci când a încercat să călărească o rază de lumină în experimentele sale mintale.

Regulile culturale îi sunt impuse neocortexului cu ajutorul creierului vechi şi mai ales al amigdalei. Toate gândurile pe care le avem declanşează alte gânduri şi unele dintre ele se referă la pericole asociate. Învăţăm, de exemplu, că încălcarea normelor culturale, chiar şi în gândurile noastre intime, poate conduce la ostracizare, pe care neocortexul o percepe ca pe o ameninţare la adresa binelui nostru. Dacă nutrim asemenea gânduri, amigdala este stimulată şi generează o stare de frică, iar aceasta, în general, duce la eliminarea gândului.

Pe de altă parte, în vis, aceste tabùuri sunt atenuate, aşa că visăm adeseori chestiuni care sunt interzise cultural, sexual sau profesional. Este ca şi cum creierul nostru şi-ar da seama că, atunci când visăm, nu suntem actori reali în lumea obiectivă. Freud a scris despre acest fenomen, dar a scos în evidenţă şi faptul că ne deghizăm asemenea gânduri periculoase, cel puţin

atunci când încercăm să le reapelăm, astfel încât creierul treaz continuă să fie protejat împotriva lor.

Relaxarea tabùurilor profesionale se dovedeşte a fi utilă pentru rezolvarea creativă a problemelor. Eu folosesc în fiecare noapte o tehnică mintală cu ajutorul căreia mă gândesc la o anumită problemă înainte să adorm. Asta îmi declanşează succesiuni de gânduri care vor continua şi în timpul somnului. Când încep să visez, mă pot gândi – *pot visa* – la soluții pentru problemă eliberat de povara limitărilor profesionale pe care o port pe timpul zilei. Pot apoi accesa aceste vise dimineața, când sunt în starea intermediară dintre vis şi veghe, numită uneori „visare lucidă".[5]

Este celebră referirea lui Freud la abilitatea de a ajunge la intuiții cu privire la psihologia persoanei prin interpretarea viselor. Există, desigur, o vastă literatură pe tema tuturor aspectelor acestei teorii, dar ideea fundamentală de obținere a unor intuiții privitoare la propria persoană prin examinarea viselor este logică. Visele sunt create de neocortexul nostru şi prin urmare substanța lor poate să fie revelatoare în privința conținutului şi a conexiunilor care se fac în el. Relaxarea constrângerilor la care este supusă gândirea noastră, constrângeri care există în starea de veghe, este de asemenea utilă pentru dezvăluirea conținutului neocortical, pe care altminteri am fi incapabili să-l accesăm direct. Este de asemenea rezonabil să concluzionăm că modelele care ajung în visele noastre reprezintă chestiuni importante pentru noi şi prin urmare sunt indicii privitoare la înțelegerea temerilor şi dorințelor noastre neîndeplinite.

Rădăcinile unui model

Aşa cum am menționat mai înainte, în anii 1980 şi 1990 am condus o echipă care a perfecționat tehnica Modelelor Markov Ascunse Ierarhic pentru recunoaşterea vorbirii umane şi înțelegerea afirmațiilor în limbaj natural. Aceste cercetări au premers sistemele comerciale răspândite de astăzi care recunosc şi înțeleg ce încercăm să le spunem (sisteme de navigație pentru automobile, cu care poți să vorbeşti, Siri pe iPhone, Google Voice Search şi multe altele). Tehnica pe care am pus-o la punct avea, în esență, toate atributele pe care le descriu în Teoria minții bazate pe recunoaşterea modelelor. Cuprindea o ierarhie de modele în care fiecare nivel superior era mai abstract conceptual decât cel de dedesubtul lui. De exemplu, în recunoaşterea vorbirii,

nivelurile includeau modele de bază ale frecvenței sunetelor la nivelul cel mai de jos, apoi foneme, apoi cuvinte şi expresii (care erau adeseori recunoscute ca şi cum ar fi fost cuvinte). O parte dintre sistemele noastre de recunoaştere a vorbirii puteau să înțeleagă semnificația comenzilor în limbaj natural, iar nivelurile mai înalte includeau structuri de felul expresiilor substantivale şi verbale. Fiecare modul de recunoaştere de model putea recunoaşte o succesiune liniară de modele de la un nivel conceptual inferior. Fiecare input avea parametri de importanță, mărime şi variabilitate a mărimii. Existau semnale „descendente" care indicau că era de aşteptat un model de nivel inferior. Voi discuta aceste cercetări mai detaliat în Capitolul 7.

În anii 2003 şi 2004, Jeff Hawkins şi Dileep George, inventatorii lui PalmPilot, au creat un model cortical ierarhic denumit memorie temporală ierarhică. Împreună cu Sandra Blakesee, specializată pe subiecte ştiințifice, Hawkins a descris acest model într-un mod elocvent în cartea lor *On Intelligence – Despre inteligență*. Hawkins demonstrează foarte convingător uniformitatea algoritmului cortical şi organizarea lui ierarhică bazată pe liste. În tre modelul prezentat în *On Intelligence* şi ceea ce prezint eu în această carte există unele deosebiri importante. Aşa cum sugerează şi numele, Hawkins pune accentul pe natura temporală (bazată pe timp) a listelor constituente. Cu alte cuvinte, direcția listelor este întotdeauna înainte în timp. Explicația lui privitoare la modul în care caracteristicile unui model, cum ar fi litera „A" tipărită, are o direcție în timp se bazează pe mişcarea ochilor. El explică faptul că noi realizăm vizualizări ale imaginilor folosind sacade – mişcări foarte rapide ale ochilor, de care nu suntem conştienți. Informația ce ajunge la neocortex este, prin urmare, nu un set bidimensional de caracteristici, ci, dimpotrivă, o listă ordonată cronologic. Deşi e adevărat că ochii noştri efectuează mişcări foarte rapide, succesiunea în care percep ei caracteristicile unui model cum ar fi litera „A" nu apare întotdeauna într-o ordine temporală sistematică. (De exemplu, sacadele ochilor nu vor înregistra întotdeauna vârful lui „A" înaintea deschiderii de la baza lui.) Mai mult, putem recunoaşte un model vizual care este prezentat cale de doar câteva zeci de milisecunde fără ca asta să însemne prea puțin timp pentru ca sacadele ochilor să-l poată scana. Este adevărat că recognizatorii de model din neocortex stochează modelul ca listă şi că lista este ordonată, dar ordinea nu exprimă în mod necesar timpul. Fireşte, adeseori aşa stau lucrurile, dar este posibil să fie vorba şi despre o ordonare spațială sau de nivel conceptual superior, aşa cum am discutat anterior.

Cea mai importantă diferență o reprezintă setul de parametri pe care i-am inclus pentru fiecare input în modulul de recunoaştere a modelului – şi în special mărimea şi parametrii de variabilitate a mărimii. În anii 1980, am încercat propriu-zis să recunoaştem vorbirea umană fără acest tip de informație. Abordarea respectivă a fost motivată de faptul că lingviştii ne-au spus că durata informației nu era deosebit de importantă. Această perspectivă este ilustrată de dicționarele care prezintă pronunția fiecărui cuvânt sub forma unui şir de foneme – de exemplu cuvântul „steep" sub forma [s] [t] [E] [p], fără vreo indicație referitoare la cât de lung e de aşteptat să fie fiecare fonem. Prin urmare, dacă creăm programe pentru recunoaşterea fonemelor şi apoi întâlnim această succesiune particulară de patru foneme (într-o formă vorbită), ar trebui să putem recunoaşte cuvântul rostit. Sistemul pe care îl construim folosind această abordare a funcționat într-o anumită măsură, dar nu suficient de bine pentru a putea face față unor atribute precum un vocabular bogat, vorbitori multipli şi cuvinte rostite continuu, fără pauze. Când am folosit tehnica Modelelor Markov Ascunse Ierarhic pentru a încorpora distribuția magnitudinilor fiecărui input, performanța a crescut imens.

Capitolul 4

NEOCORTEXUL BIOLOGIC

Pentru că lucrurile importante stau în cutii, ai craniul pentru creier, un etù din plastic pentru pieptene şi portofelul pentru bani.

— George Costanza, în episodul „Vizorul invers" din *Seinfeld*

Acum, pentru prima oară, observăm creierul în funcțiune într-o manieră globală, cu o asemenea claritate, încât ar trebui să fim în stare să descoperim programele complète din spatele acestor magnifice puteri.

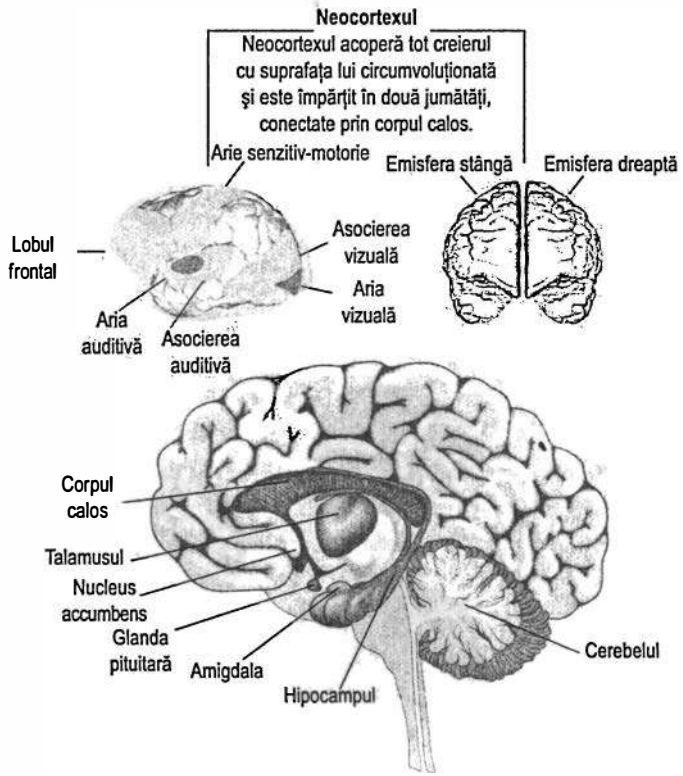
— J.G. Taylor, B. Horwitz şi K.J. Friston

Mintea, pe scurt, lucrează cu datele pe care le primeşte, foarte asemănător cu sculptorul care lucrează cu blocul lui de piatră. Într-un sens, statuia se afla în el de când lumea. Dar mai existau o mie de altele împreună cu ea şi numai sculptorului trebuie să i se mulțumească pentru faptul că a degajat-o pe aceasta dintre toate celelalte. Tot astfel, lumile fiecăruia dintre noi, oricât de diferite ar fi punctele noastre de vedere despre ele, sunt înglobate în haosul primordial al senzațiilor, care a dat *materia* gândului fiecăruia dintre noi atât de fără deosebire. Putem, dacă vrem, prin raționamente, să descompunem lucrurile până la acea stare de întuneric şi continuitate fără cusături a spațiului şi a norilor în mişcare de atomi aglomerați pe care ştiința o numeşte singura lume reală. Cu toate acestea, lumea pe care *noi* o simțim şi în care trăim va fi aceea pe care străbunii noştri şi noi, prin loviturile încet cumulative ale alegerilor, am eliberat-o, asemenea sculptorilor, aruncând pur şi simplu unele părți din materia ce ne-a fost dată. Alți sculptori, alte statui, din aceeaşi piatră! Alte minți, alte lumi, din acelaşi monoton şi inexpresiv haos! Lumea mea nu este decât una dintr-un milion de altele întipărite, la fel de reale pentru cei care ar putea să le abstractizeze. Cât de diferite trebuie să fie lumile în conştiința unei furnici, a peştelui-cuțit sau a crabului!

— William James

Este inteligența scopul sau măcar unul dintre scopurile evoluției biologice? Steven Pinker scrie: „Suntem șovini în privința creierului nostru, gândind că el este scopul evoluției"[1] și continuă susținând că „Asta nu are niciun sens... Selecția naturală nu face nici pe departe ceva care să indice că se străduiește pentru inteligență. Procesul e propulsat de deosebirile ratelor de supraviețuire și reproducere ale organismelor care se înmulțesc într-un anumit mediu. Cu timpul, organismele capătă particularități care le adaptează pentru supraviețuire și reproducere în acel mediu. Punct. Nimic nu le trage în vreo altă direcție decât succesul acolo și atunci". Pinker concluzionează că „viața e o tufă bogat ramificată, nu o balanță sau o scară, iar organismele vii se află în vârfurile ramurilor, nu pe treptele inferioare".

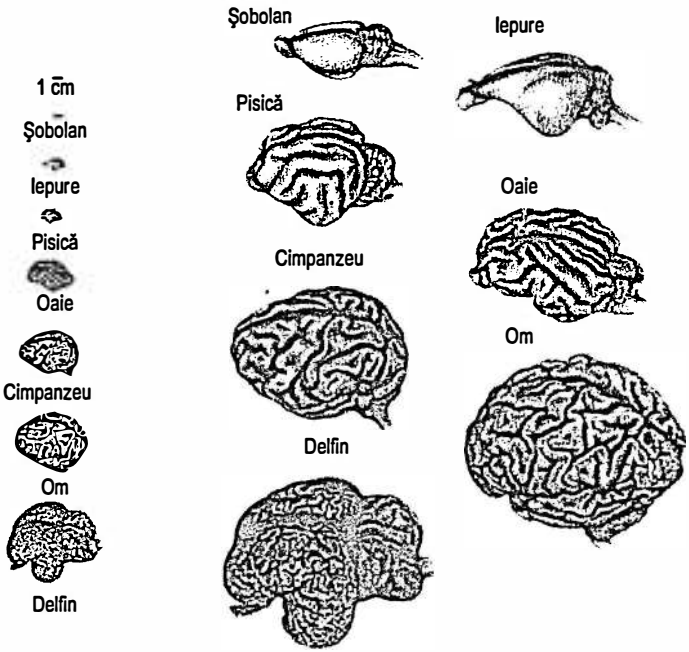
În privința creierului uman, Pinker se întreabă dacă „beneficiile depășesc costurile". Dintre costuri, el îl citează pe acela că „creierul [este]



Dispunerea fizică a regiunilor-cheie ale creierului

voluminos. În bazinul femeii, capul supradimensionat al copilului abia dacă poate să aibă loc. Acest compromis de proiectare ucide multe mame la naştere şi impune un mers pivotant care le face pe femei să aibă o mişcare bipedă mai puţin eficientă din punct de vedere biomecanic decât bărbaţii. De asemenea, un cap greu care se susţine pe gât ne face mai vulnerabili la accidente fatale, cum ar fi căzăturile". Pinker continuă cu enumerarea altor neajunsuri, inclusiv consumul de energie al creierului, timpul mare de reacţie şi procesul îndelungat de învăţare.

Deşi toate aceste afirmaţii sunt exacte la prima vedere (şi cu toate că multe dintre prietenele mele merg pe jos mai bine decât mine), lui Pinker îi scapă, aici, aspectul de ansamblu. Este foarte adevărat că, din punct de vedere biologic, evoluţia nu are o direcţie specifică. Ea reprezintă o metodă de căutare care încarcă temeinic „ramurile bogat ramificate" ale naturii. La fel de adevărat este şi că schimbările evolutive nu se îndreaptă *neapărat* în direcţia unei inteligenţe sporite – se îndreaptă în *toate* direcţiile. Există



Neocortexul la diferite mamifere

multe exemple de ființe pline de succes care au rămas complet neschimbate timp de milioane de ani. (Aligatorii, de exemplu, sunt neschimbați de 200 de milioane de ani şi numeroase microorganisme – de mult mai mult timp.) Dar, în cursul încărcării temeinice a nenumăratelor ramuri evolutive, una dintre direcțiile în care se înaintează este o inteligență mai mare. Acesta e aspectul relevant pentru scopul acestei discuții.

Să presupunem că, într-un borcan, avem un gaz albastru. Când scoatem capacul, nu există niciun mesaj care să fie transmis tuturor moleculelor de gaz şi care să le spună: „Hei, băieți, borcanul nu mai are capacul pus; hai să ne ducem spre gura lui şi să ieşim". Moleculele pur şi simplu fac ceea ce fac dintotdeauna, adică se mişcă în toate părțile, aparent fără o direcție anume. Dar, în timp ce fac asta, unele dintre ele, aflate aproape de gura borcanului, chiar ies din el şi, cu timpul, majoritatea moleculelor le urmează exemplul. Din momentul în care a nimerit un mecanism neuronal capabil de învățare ierarhizată, evoluția biologică a constatat că acesta e imens de util pentru unicul obiectiv al evoluției, care este supraviețuirea. Beneficiul posesiei unui neocortex a devenit acut în momentul când circumstanțele în rapidă schimbare au necesitat învățarea rapidă. Specii de toate felurile – plante şi animale – pot învăța să se adapteze la circumstanțe în schimbare, dar, în lipsa neocortexului, trebuie să folosească procesul de evoluție genetică. Poate fi nevoie de foarte multe generații – mii de ani – pentru ca o specie lipsită de neocortex să învețe noi comportamente semnificative (sau, în cazul plantelor, noi strategii de adaptare). Avantajul izbitor pentru supraviețuire al neocortexului a constat în faptul că acesta putea să învețe în decurs de câteva zile. Dacă o specie se confruntă cu o schimbare dramatică a condițiilor de viață şi un membru al speciei inventează, descoperă sau pur şi simplu nimereşte peste un mijloc de adaptare la această schimbare (aceste trei metode fiind, toate, variații ale inovației), alți indivizi observă, învață şi copiază metoda, iar aceasta se răspândeşte rapid, viral, la nivelul întregii populații. Evenimentul extincției cataclismice din Cretacic-Paleogen, din urmă cu aproximativ 65 de milioane de ani, a condus la dispariția rapidă a multor specii care nu posedau neocortex şi deci nu s-au putut adapta suficient de rapid la mediul brusc modificat. Acesta a fost punctul de turnură pentru mamiferele capabile de neocortex în acțiunea de a-şi asigura nişa ecologică. În acest fel, evoluția biologică a constatat că proprietatea de învățare ierarhizată a neocortexului era atât de valoroasă, încât această regiune a creierului a continuat să îşi extindă dimensiunile până când practic a învăluit creierul lui *Homo sapiens*.

Descoperirile din domeniul neuroştiinţelor au stabilit într-un mod convingător rolul-cheie jucat de capacităţile ierarhice ale neocortexului şi în acelaşi timp au adus dovezi în sprijinul Teoriei minţii bazate pe recunoaşterea modelelor. Dovezile constau în numeroase observaţii şi analize, dintre care voi trece o parte în revistă aici. Psihologul canadian Donald O. Hebb (1904–1985) a făcut o primă încercare de a explica baza neurologică a învăţării. În anul 1949, el a descris un mecanism graţie căruia neuronii se modifică fiziologic pe baza experienţei proprii, asigurând astfel temelia pentru învăţare şi pentru plasticitatea creierului: „Să presupunem că persistenţa repetiţiei unei activităţi reverberative (sau «urma») tinde să inducă modificări celulare durabile care îi sporesc stabilitatea... Când axonul unei celule *A* se află suficient de aproape încât să excite o celulă *B* şi când, în mod repetat sau persistent, celula *A* participă la excitarea celulei *B*, are loc un proces de creştere ori o modificare metabolică în una sau în ambele celule, astfel încât eficienţa lui *A*, care este una dintre celulele ce excită celula *B*, sporeşte".[2] Această teorie a fost enunţată sub forma „Celulele care excită împreună sunt conectate împreună" şi a devenit cunoscută sub numele de Învăţare hebbiană. Aspecte ale teoriei lui Hebb au fost confirmate: este clar că ansamblurile creierului pot să creeze noi conexiuni şi să le întărească pe baza propriei lor activităţi. Practic, putem vedea că neuronii îşi dezvoltă asemenea conexiuni în scanările cerebrale. „Reţelele neuronale" artificiale se bazează pe modelul învăţării neuronale al lui Hebb.

Ipoteza centrală a teoriei lui Hebb este că unitatea fundamentală de învăţare la nivelul neocortexului o reprezintă neuronul. Teoria minţii bazate pe recunoaşterea modelelor pe care o enunţ în această carte se bazează pe o unitate fundamentală diferită: nu neuronul însuşi, ci mai curând un ansamblu de neuroni, pe care îl estimez la circa o sută. Conectarea şi tăria sinapselor la nivelul fiecărei unităţi sunt relativ stabile şi determinate genetic – respectiv, organizarea la nivelul fiecărui modul de recunoaştere a modelelor este determinată de proiectul genetic. Învăţarea se realizează prin crearea de conexiuni între aceste unităţi şi nu în interiorul lor, dar şi, probabil, prin tăriile sinaptice ale conexiunilor din cadrul unităţilor.

În sprijinul ipotezei că modulul de bază pentru învăţare este unul compus din zeci de neuroni vine specialistul elveţian în neuroştiinţă Henry Markram (născut în 1962), căruia îi descriu ambiţiosul Proiect Blue Brain de simulare a întregului creier uman în Capitolul 7. Într-o lucrare din 2011, el dezvăluie faptul că, atunci când a scanat şi analizat neuroni de mamifere,

potrivit propriei formulări „căutam dovezi ale ansamblurilor hebbiene la nivelul cel mai elementar al cortexului". Ce a descoperit în schimb, scrie el, au fost „ansambluri evazive [ale căror] conectivități și greutăți sinaptice sunt extrem de previzibile și restricționate". Markram concluzionează că „aceste descoperiri implică faptul că experiența nu poate modela cu ușurință conexiunile sinaptice ale ansamblurilor" și speculează că „pentru percepție, [ele] au nevoie de blocuri de cunoaștere înnăscute, de genul pieselor de Lego, și că achiziția de amintiri implică o combinare a acestor blocuri în construcții complexe". Markram continuă:

> De decenii întregi a fost comunicată existența ansamblurilor neuronale funcționale, dar dovezi directe ale existenței aglomerărilor de neuroni conectați sinaptic... au lipsit. ...De vreme ce aceste ansambluri sunt toate similare ca topologie și greutate sinaptică, nefiind modelate de vreo experiență specifică, noi considerăm că sunt ansambluri înnăscute... Experiența joacă un rol minor în determinarea conexiunilor sinaptice și a greutății lor în cadrul acestor ansambluri... Studiul nostru a condus la descoperirea de dovezi [ale existenței] ansamblurilor înnăscute de genul pieselor Lego, compuse din câteva zeci de neuroni... Conexiunile dintre ansambluri pot să le combine pe acestea în superansambluri la nivelul unui strat neocortical, apoi în ansambluri de ordin superior într-o coloană corticală, ba chiar în superansambluri într-o regiune a creierului și, în cele din urmă, într-un ansamblu de ordinul cel mai înalt, reprezentat de întregul creier. ...Achiziția de amintiri este foarte asemănătoare construirii cu piese Lego. Fiecare ansamblu e echivalent cu o piesă Lego care conține cunoștințe elementare înnăscute referitoare la modul de procesare, de percepție și de răspuns la lume. ...Când se reunesc, mai multe blocuri diferite pot, prin urmare, să formeze o combinație unică a acestor percepții înnăscute, care reprezintă cunoștințele și experiențele specifice ale individului.[3]

„Piesele Lego" pe care le propune Markram sunt pe deplin compatibile cu modulele de recunoaștere a modelelor pe care le-am descris eu. Într-o comunicare prin e-mail, Markram a descris aceste „piese Lego" ca fiind „conținut împărtășit și cunoștințe înnăscute".[4] Eu aș afirma că scopul acestor module este să recunoască modele, să le memoreze și să le anticipeze pe baza unor modele parțiale. A se remarca faptul că estimarea lui

Markram potrivit căreia fiecare modul ar conține „câteva zeci de neuroni" se bazează numai pe stratul V al neocortexului. Stratul V este într-adevăr bogat în neuroni, dar, raportat la numărul obișnuit de neuroni estimați pe stratul VI, asta s-ar traduce printr-un ordin de mărime de aproximativ 100 de neuroni per modul, cifră aflată în concordanță cu estimările mele.

Interconexiunile consistente și aparentul caracter modular al neocortexului au fost observate de ani de zile, dar studiul descris anterior e primul care demonstrează stabilitatea acestor module atunci când creierul își derulează procesele dinamice.

Un alt studiu recent, realizat de data asta la General Hospital din Massachusetts, finanțat de National Institutes of Health și de National Science Foundation și dat publicității într-o ediție din luna martie a anului 2012 a revistei *Science*, vorbește de asemenea despre o structură de conexiuni la nivelul întregului neocortex.[5] Articolul arată că, la nivelul neocortexului, conexiunile respectă un model de grilă similar celui al străzilor ordonate dintr-un oraș: „În esență, structura de ansamblu a creierului ajunge să se asemene cu Manhattan-ul, în care ai o hartă 2-D a străzilor și o axă, un lift, care intră în a treia dimensiune", a scris Van J. Wedeen, specialist în neuroștiință și fizician la Harvard, șeful studiului.

Într-un podcast al revistei *Science*, Wedeen a descris semnificațiile cercetărilor: „A fost un studiu asupra structurii tridimensionale a căilor situate la nivelul creierului. Când oamenii de știință s-au gândit la căile din creier în ultima sută de ani, imaginea tipică sau modelul care le-a venit în minte a fost acela al unor căi asemănătoare cu spaghetele dintr-un bol – trasee separate, cu modele spațiale prea puțin legate unele de altele. Folosind rezonanța magnetică, am putut investiga experimental această chestiune. Și am descoperit că, în loc să existe căi aranjate întâmplător sau independente, toate traseele din creier corespund unei structuri extraordinar de simple. În esență, ansamblul seamănă cu un cub. Căile se întind pe cele trei direcții perpendiculare și, pe fiecare dintre aceste trei direcții, sunt perfect paralele între ele și sunt aranjate în serii. Deci, în locul unui bol de spaghete independente, observăm că, că la nivelul creierului, conectivitatea este, într-un sens, o unică structură coerentă".

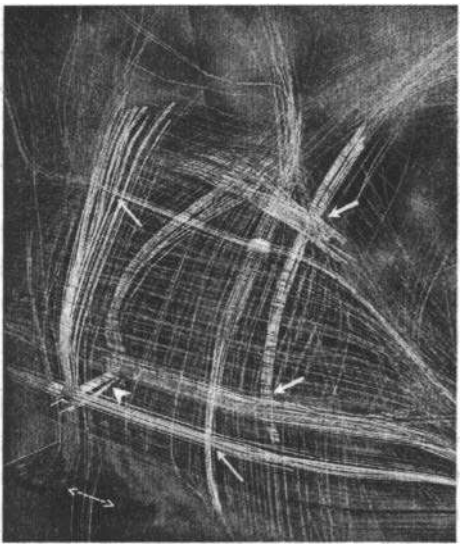
În vreme ce studiul lui Markram prezintă un modul de neuroni care se repetă la nivelul întregului neocortex, studiul lui Wedeen demonstrează existența unui model remarcabil de ordonat de conexiuni între module. Creierul pornește de la un număr foarte mare de „conexiuni în așteptare", la care se

pot lega modulele de recunoaştere a modelelor. Prin urmare, dacă un anumit modul vrea să se conecteze cu altul, nu trebuie să aibă un axon din unul şi o dendrită din celălalt care să acopere distanța fizică dintre ele. Poate, pur şi simplu, să acceseze una dintre aceste conexiuni axonice în aşteptare şi să se lege automat la capetele fibrei. Aşa cum scriu şi Wedeen şi colaboratorii săi: „Căile creierului respectă un plan de bază stabilit de... embriogeneza timpurie. Prin urmare, căile creierului ajuns la maturitate reflectă aceste trei gradiente primordiale deformate fizic de dezvoltare". Cu alte cuvinte, pe măsură ce învățăm şi avem experiențe, modulele de recunoaştere a modelelor din neocortex se conectează la aceste legături preexistente care au fost create pe când eram embrioni.

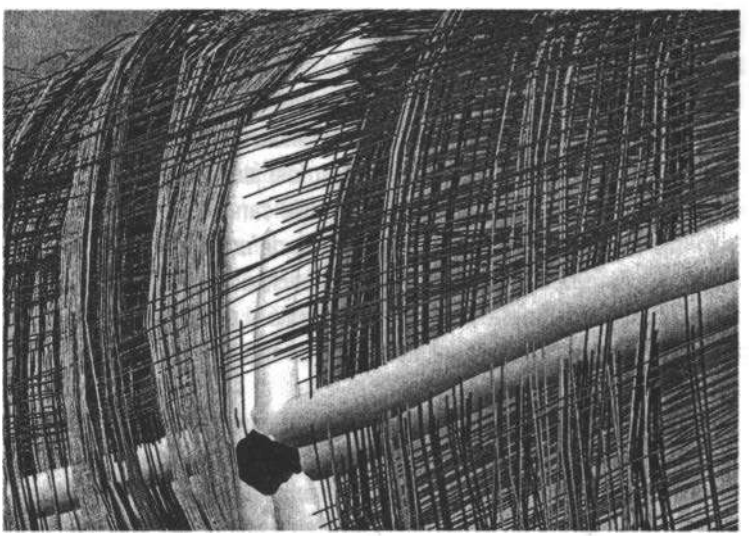
Există un cip electronic, numit *Field Programmable Gate Array (FPGA)* – serie de porți de câmp programabile –, care se bazează pe un principiu asemănător. Cipul conține milioane de module care implementează funcții logice şi creează conexiuni în aşteptare. În momentul utilizării, conexiunile respective nu sunt nici activate, nici dezactivate (prin semnale electronice) în vederea implementării unei anumite capacități.

La nivelul neocortexului, aceste conexiuni la mare distanță care rămân nefolosite sunt eventual tăiate. Este unul dintre motivele pentru care adaptarea unei regiuni învecinate a neocortexului în vederea compensării uneia lezate nu este la fel de eficientă ca regiunea originală. Potrivit studiului lui Wedeen, conexiunile inițiale sunt extrem de ordonate şi repetitive, exact la fel ca modulele propriu-zise, iar modelul lor în grilă este folosit pentru „ghidarea conectivității" la nivelul neocortexului. Acest model a fost identificat la toate creierele de primate şi de oameni care au fost studiate şi s-a dovedit a fi evident la nivelul neocortexului de la regiunile care se ocupă de modelele senzoriale timpurii, până la cele responsabile de emoțiile de nivel superior. Articolul lui Wedeen din revista *Science* concluzionează că „structura de grilă a căilor cerebrale este universală, coerentă şi continuă pe cele trei axe principale de dezvoltare". Acest fapt indică un algoritm comun pentru toate funcțiile neocortexului.

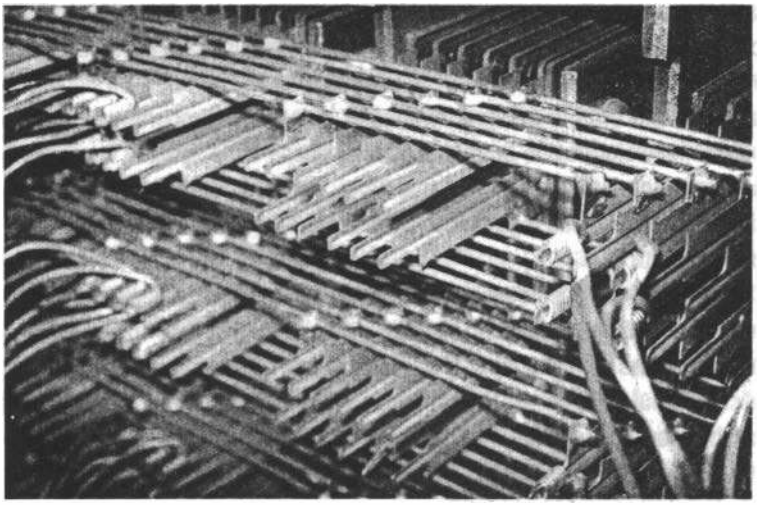
Este cunoscut de mult faptul că măcar anumite regiuni ale neocortexului sunt ierarhice. Regiunea cel mai bine studiată este cortexul vizual, separat în arii denumite V1, V2 şi MT (cunoscută de asemenea şi ca V5). Pe măsură ce înaintăm în arii superioare din această regiune („superioare" în sensul procesării conceptuale, nu fizic, deoarece neocortexul are mereu o grosime egală cu un singur recognizator de modele), proprietățile care pot fi recunoscute



Structura extraordinar de regulată a conexiunilor inițiale de la nivelul neocortexului, descoperită în cadrul unui studiu al National Institutes of Health



O altă vedere a structurii regulate de tip grilă a conexiunilor de la nivelul neocortexului



Structura de tip grilă descoperită la nivelul neocortexului este remarcabil de asemănătoare cu ceea ce se numeşte comutator încrucişat, care e folosit în circuitele integrate şi în plăcile de circuite

devin tot mai abstracte. Aria V1 recunoaşte muchiile elementare şi formele primitive. Aria V2 recunoaşte contururile, diferențele dintre imaginile prezentate de fiecare ochi, orientarea spațială şi posibilitatea ca o porțiune din imagine să facă parte din obiect sau din fundal ori ba.[6] Regiunile de nivel superior ale neocortexului recunosc concepte precum identitatea obiectelor şi a chipurilor şi mişcările lor. De asemenea, este de multă vreme cunoscut faptul că orice comunicare prin intermediul acestei ierarhii se face atât în sens ascendent, cât şi descendent şi că semnalele pot fi atât excitatoare, cât şi inhibitoare. Tomaso Poggio (născut în 1947), specialist în neuroştiință la MIT, a studiat în detaliu vederea la nivelul creierului uman, iar cercetările sale din ultimii treizeci şi cinci de ani au fost esențiale pentru localizarea învățării ierarhice şi pentru recunoaşterea modelelor la nivelurile „timpurii" (cele mai reduse conceptual) ale neocortexului vizual.[7]

Ceea ce înțelegem noi cu privire la nivelurile ierarhice inferioare ale neocortexului vizual corespunde Teoriei minții bazate pe recunoaşterea modelelor pe care am descris-o în capitolul anterior, iar observațiile privitoare la natura ierarhică a procesării neocorticale s-au extins recent cu mult dincolo de aceste niveluri. Daniel J. Felleman, profesor de neurobiologie

la Universitatea Texas, şi colegii săi au identificat „organizarea ierarhică a cortexului cerebral... [în] 25 de arii ale neocortexului", care includ atât ariile vizuale, cât şi pe cele de nivel superior care combină modelele primite de la mai multe organe de simţ. Avansând în ierarhia neocortexului, procesarea modelelor devine tot mai abstractă, ocupă zone spaţiale tot mai largi şi presupune intervale de timp mai lungi. Pentru fiecare conexiune, ei au descoperit în ierarhie legături de comunicare atât în sens ascendent, cât şi descendent.[8]

Cercetările recente ne permit să ne lărgim substanţial aceste observaţii înspre regiuni care depăşesc cu mult cortexul vizual şi chiar înspre ariile de asociere, care combină input-ul primit de la mai multe simţuri. Un studiu publicat în anul 2008 de Uri Hasson, profesor de psihologie la Princeton, şi de colaboratorii săi demonstrează că fenomenele observate la nivelul cortexului vizual se produc în numeroase arii ale neocortexului: „Este bine stabilit faptul că neuronii de pe căile corticale vizuale au câmpuri receptoare spaţiale din ce în ce mai mari. Este un principiu de organizare elementar al sistemului vizual... Evenimentele din lumea reală apar nu doar în regiuni ample ale spaţiului, ci şi în perioade extinse de timp. Prin urmare, am formulat ipoteza că ar trebui să existe o ierarhie analoagă celei descoperite pentru mărimea câmpurilor receptoare spaţiale şi pentru caracteristicile de răspuns temporal ale diferitelor regiuni ale creierului". Şi exact asta au descoperit, ceea ce le-a permis să concluzioneze că, „similar ierarhiei corticale cunoscute a câmpurilor receptoare spaţiale, în creierul omului există o ierarhie a ferestrelor receptoare temporale progresiv crescătoare".[9]

Cel mai convingător argument privitor la universalitatea procesării la nivelul neocortexului îl reprezintă dovada omniprezentă a plasticităţii (nu doar învăţare, ci şi interschimbabilitate); cu alte cuvinte, o regiune e capabilă să facă munca altor regiuni, fapt ce implică existenţa unui algoritm comun la nivelul întregului neocortex. O parte foarte importantă a cercetărilor din domeniul neuroştiinţei s-a concentrat asupra stabilirii regiunilor din neocortex care au legătură cu anumite tipuri de modele.

Tehnica clasică de determinare a acestor legături a luat în considerare creierele care au suferit leziuni în urma unor traumatisme sau atacuri cerebrale şi asocierea funcţiilor pierdute cu regiunile specifice lezate. De exemplu, atunci când observăm că o persoană care a suferit o leziune de dată recentă în regiunea girusului fusiform are dificultăţi în recunoaşterea chipurilor, dar îşi poate identifica semenii după voce şi modele de limbaj,

putem emite ipoteza că respectiva regiune are legătură cu recunoașterea chipurilor. Ipoteza fundamentală a fost aceea că fiecare dintre regiunile cu pricina este concepută să recunoască și să proceseze un anumit tip de model. Anumite regiuni fizice au fost asociate cu anumite tipuri de modele, deoarece, în condiții normale, așa circulă informația. Dar, când fluxul normal de informație este tulburat dintr-un motiv oarecare, o altă regiune a neocortexului se vădește a fi capabilă să intervină și să preia funcția.

Plasticitatea a fost larg observată de neurologi, care au sesizat că pacienții cu leziuni cerebrale survenite în urma unui traumatism sau a unui atac cerebral pot reînvăța aceleași abilități într-o altă regiune a neocortexului. Probabil, exemplul cel mai dramatic de plasticitate este de găsit într-un studiu din anul 2011 al specialistului american în neuroștiință Marina Bedny și al colaboratorilor săi. Studiul vizează situația cortexului vizual la persoanele oarbe congenital. Opinia generală era că straturile timpurii ale cortexului vizual, cum ar fi V1 și V2, se ocupă în mod inerent de modelele inferioare (ca de pildă muchiile și curbele), în vreme ce cortexul frontal (acea nouă regiune evolutivă a cortexului pe care o avem în spatele frunților noastre unic de largi) se ocupă în mod inerent de modelele mai complexe și mai subtile ale limbajului și de conceptele abstracte. Dar, așa cum au descoperit Bedny și colaboratorii săi, „se crede că omul posedă la nivelul lobilor frontali și temporali stâng și drept regiuni cerebrale evoluate, capabile în mod unic de procesarea limbajului. Pe de altă parte însă, persoanele oarbe congenital activează și cortexul vizual pentru anumite funcții verbale. Noi aducem dovezi că această activitate a cortexului vizual reflectă, de fapt, procesarea limbajului. Am descoperit că, la persoanele oarbe congenital, cortexul vizual stâng se comportă similar regiunilor clasice ale limbajului... Concluzionăm că regiunile creierului care se consideră că au evoluat în scopul vederii pot adopta procesarea limbajului ca rezultat al unei experiențe timpurii".[10] Gândește-te numai, cititorule, la implicațiile acestui studiu: înseamnă că regiuni ale neocortexului care sunt relativ depărtate sub raport fizic și care au fost considerate și foarte diferite din punct de vedere conceptual (sugestii vizuale primitive *versus* concepte abstracte de limbaj) folosesc în esență același algoritm. Regiunile care procesează aceste tipuri disparate de modele se pot substitui reciproc.

Daniel E. Feldman, specialist în neuroștiință la Universitatea California, Berkeley, a scris o cuprinzătoare recenzie pe tema a ceea ce el a numit „mecanismele sinaptice pentru plasticitatea neocortexului" și a descoperit dovezi referitoare la acest tip de plasticitate la nivelul întregului neocortex. El

notează că „plasticitatea îi permite creierului să învețe și să-și reamintească modele din lumea senzorială, să rafineze mișcări... și să-și recapete funcția după leziuni". El adaugă și că această plasticitate este făcută posibilă de „modificări structurale, inclusiv compunerea, îndepărtarea și remodelarea morfologică a sinapselor corticale și a crestelor dendritice".[11]

Un alt exemplu uluitor de plasticitate neocorticală (și, prin urmare, venind în sprijinul uniformității algoritmului neocortical) a fost demonstrat recent de oamenii de știință de la Universitatea California, Berkeley. Ei au montat serii de microelectrozi pentru culegerea de semnale cerebrale dintr-o regiune a cortexului motor al șoarecilor care controlează mișcările mustăților. Cercetătorii au conceput experimentul astfel încât șoarecii să obțină o recompensă dacă își controlau acești neuroni pentru a genera un anumit model mintal fără să-și miște mustățile. Modelul necesar pentru obținerea recompensei presupunea o sarcină mintală pe care neuronii frontali ai șoarecilor nu ar fi îndeplinit-o în mod normal. Însă șoarecii s-au dovedit capabili să îndeplinească această sarcină mintală gândind în esență cu neuronii motori pe care îi decuplau mintal de la controlul mișcărilor motorii.[12] Concluzia este că neocortexul motor, regiune a neocortexului responsabilă de coordonarea mișcărilor musculare, folosește și el un algoritm neocortical standard.

Pe de altă parte, există mai multe motive pentru care o abilitate sau un domeniu de cunoaștere reînvățat cu ajutorul unei noi arii a neocortexului, care o înlocuia pe una lezată, nu este la fel de bună ca și cea originală. În primul rând, deoarece a fost nevoie de o viață întreagă pentru învățarea și perfecționarea unei anumite abilități, reînvățarea într-o altă arie a neocortexului nu poate genera imediat aceleași rezultate. Mai important, acea altă arie a neocortexului nu a stat degeaba, pe post de rezervă pentru regiunea lezată. Și ea a îndeplinit funcții vitale, așa că va ezita să renunțe la modelele sale neocorticale pentru a compensa regiunea lezată. Ea poate începe prin a elibera o parte dintre còpiile redundante ale modelelor sale, dar, prin aceasta, va degrada în mod subtil abilitățile existente și nu va elibera atât de mult spațiu cortical cât folosiseră inițial abilitățile care sunt reînvățate acum.

Mai există și un al treilea motiv pentru care plasticitatea are limitele ei. De vreme ce la majoritatea oamenilor tipurile de modele se transmit prin regiuni specifice (de exemplu, chipurile sunt procesate de girusul fusiform), aceste regiuni au ajuns să fie optimizate (de evoluția biologică) pentru respectivele tipuri de modele. După cum voi arăta în Capitolul 7, am descoperit

aceleași rezultate în creațiile noastre neocorticale digitale. Am putea recunoaște vorbirea cu sistemele noastre de recunoaștere a caracterelor și viceversa, dar sistemele de vorbire au fost optimizate pentru vorbire și, în mod similar, sistemele de recunoaștere a caracterelor au fost optimizate pentru caracterele tipărite, deci ar exista o anumită diminuare a performanței dacă l-am substitui pe unul cu celălalt. Am folosit practic algoritmi evolutivi (genetici) pentru realizarea acestei optimizări, o simulare a ceea ce biologia face în mod natural. Dat fiind faptul că chipurile au fost transmise prin girusul fusiform la majoritatea oamenilor timp de sute de mii de ani (sau mai mult), evoluția biologică a avut timpul necesar pentru a dezvolta o capacitate favorabilă procesării unor asemenea modele în acea regiune. Girusul fusiform folosește aceiași algoritmi de bază, dar este orientat către chipuri. Așa cum a scris specialistul olandez în neuroștiință Randal Koene, „neocortexul este foarte uniform, fiecare coloană sau minicoloană putând să facă, în principiu, tot ceea ce face o alta".[13]

Cercetări substanțiale susțin observația că modulele de recunoaștere a modelelor se conectează pe baza modelelor la care sunt expuse. De exemplu, specialistul în neuroștiință Yi Zuo și colegii ei au urmărit felul cum noi „spini dendritici" formau conexiuni între celulele nervoase în timp ce șoarecii învățau o nouă abilitate (să întindă laba printr-o fantă ca să apuce o sămânță).[14] Cercetătorii de la Institutul Salk au descoperit că această autoconectare fundamentală a modulelor neocortexului este controlată, aparent, de numai câteva gene. Aceste gene și această metodă de autoconectare sunt de asemenea uniforme la nivelul neocortexului.[15]

Multe alte studii documentează aceste atribute ale neocortexului, dar să rezumăm ce putem observa din literatura neuroștiinței și din propriile experimente mintale. Unitatea de bază a neocortexului este un modul de neuroni al căror număr îl estimez la circa o sută. Aceste module sunt întrețesute în fiecare coloană a neocortexului, astfel încât un modul în parte nu este vizibil distinct. Modelul conexiunilor și tăriilor sinaptice din cadrul fiecărui modul este relativ stabil. Tocmai conexiunile și tăriile sinaptice dintre module sunt cele care reprezintă învățarea.

La nivelul neocortexului există ceva de ordinul unui cvadrilion (10^{15}) de conexiuni, dar numai în jur de 25 de biți de informație creativă la nivelul genomului (după compresia fără pierdere)[16], astfel încât conexiunile propriu-zise nu pot fi în niciun caz predeterminate genetic. Este posibil ca o parte din această învățare să fie produsul interogării creierului vechi de către

neocortex, dar şi asta ar reprezenta doar un volum relativ mic de informaţie. Conexiunile dintre module sunt generate în ansamblul lor în urma experienţei (prin cultivare, mai degrabă decât de la natură).

Creierul nu are suficientă flexibilitate, astfel încât fiecare modul de recunoaştere a modelelor se poate conecta pur şi simplu cu orice alt modul (aşa cum ne putem programa cu uşurinţă computerele sau cum ne putem lega la web). Trebuie să fie realizată o conexiune fizică propriu-zisă, formată dintr-un axon legat de o dendrită. Toţi ne începem viaţa cu un vast stoc de posibile conexiuni neuronale. După cum demonstrează studiul lui Wedeen, aceste conexiuni sunt organizate de o manieră hiperrepetitivă şi ordonată. Conexiunea terminală cu aceşti axoni în aşteptare se produce pe baza modelelor pe care le-a recunoscut fiecare recognizator de modele din neocortex. Conexiunile neutilizate sunt în cele din urmă tăiate. Aceste conexiuni sunt construite ierarhic, reflectând ordinea ierarhică naturală a realităţii. Aceasta este punctul-forte cheie al neocortexului.

Algoritmul de bază al modulelor de recunoaştere a modelelor din neocortex este acelaşi pentru întregul neocortex, de la modulele de „nivel redus", care se ocupă de cele mai elementare modele senzoriale, până la modulele de „nivel superior", care recunosc cele mai abstracte concepte. Vasta evidenţă a plasticităţii şi interschimbabilităţii regiunilor neocortexului este mărturia acestei importante observaţii. Există o anumită optimizare a regiunilor care se ocupă cu anumite tipuri de modele, dar acest efect este de ordin secund – algoritmul fundamental e universal.

Semnalele se transmit ascendent şi descendent în ierarhia conceptuală. Un semnal cu sens ascendent înseamnă „Am detectat un model". Un semnal cu sens descendent înseamnă „Mă aştept să apară modelul tău" şi este, în esenţă, o predicţie. Atât semnalele ascendente, cât şi cele descendente pot fi ori excitatoare ori inhibitoare.

Fiecare model e poziţionat el însuşi într-o anumită ordine, deci este greu de inversat. Chiar dacă un model pare să aibă aspecte multidimensionale, el e reprezentat de o succesiune unidimensională de modele de nivel inferior. Modelul este o succesiune ordonată de alte modele, astfel încât fiecare recognizator este în mod inerent recursiv. Pot să existe numeroase nivele ierarhice.

În modelele pe care le învăţăm se află o extraordinară redundanţă, mai ales în cele importante. Recunoaşterea modelelor (cum ar fi obiectele şi chipurile obişnuite) apelează la aceleaşi mecanisme ca amintirile, care

sunt simple modele pe care le-am învățat. Ele sunt de asemenea stocate ca succesiuni de modele – adică sunt în esență povești. Acest mecanism este folosit și pentru învățarea și realizarea mișcării fizice în lume. Redundanța modelelor ne permite să recunoaștem obiecte, oameni și idei chiar și atunci când apar variațiuni, iar contextele sunt diferite. De asemenea, dimensiunea și variabilitatea mărimii parametrilor îi permit neocortexului să codifice variația de magnitudine comparativ cu diferite unități de mărime (durata, în cazul sunetului). Un mod de a codifica acești parametri de magnitudine îl constituie modelele multiple cu numere diferite de input-uri repetate. Așa, de exemplu, pot să existe modele pentru cuvântul rostit „steep" cu numere diferite de vocale lungi repetate [E], fiecare cu parametrul de importanță stabilit la un nivel moderat, care să indice faptul că repetarea lui [E] este variabilă. Această abordare nu e matematic echivalentă cu faptul de a avea parametri de mărime expliciți și nu dă rezultate la fel de bune în practică, însă reprezintă o abordare a codificării magnitudinii. Cea mai elocventă dovadă pe care o avem în privința acestor parametri o reprezintă faptul că ei sunt necesari sistemelor noastre de inteligență artificială pentru obținerea unor niveluri de acuratețe apropiate de cele umane.

Rezumatul de mai înainte constituie concluziile pe care le-am tras din experimentele despre care am discutat anterior și din experimentele mintale despre care de asemenea am discutat. Susțin în continuare că modelul pe care l-am prezentat este singurul care poate satisface toate constrângerile impuse de cercetările și experimentele noastre mintale.

În final, mai există o mărturie care poate fi coroborată. Tehnicile pe care le-am perfecționat pe parcursul ultimelor decenii în domeniul inteligenței artificiale pentru recunoașterea și procesarea inteligentă a fenomenelor din lumea reală (cum ar fi limbajul uman vorbit și cel scris) și pentru înțelegerea documentelor în limbaj natural se dovedesc a fi similare din punct de vedere matematic modelului pe care l-am prezentat mai înainte. Sunt de asemenea exemple ale Teoriei minții bazate pe recunoașterea modelelor. Domeniul inteligenței artificiale nu a încercat în mod explicit să copieze creierul, dar, cu toate acestea, a ajuns la tehnici în esență echivalente.

Capitolul 5

CREIERUL VECHI

Am un creier vechi, dar o memorie extraordinară.

— Al Lewis

Iată-ne în mijlocul acestei noi lumi cu creierul nostru primitiv acordat la viața simplă de cavernă, dar cu forțe extraordinare la dispoziția noastră, pe care suntem destul de deștepți să le descătușăm, însă cărora nu le putem înțelege consecințele.

— Albert Szent-Györgyi

Creierul nostru vechi – cel pe care l-am avut înainte de a fi mamifere – nu a dispărut. Continuă, firește, să asigure o mare parte din motivația noastră pentru căutarea recompenselor şi evitarea pericolelor. Aceste scopuri sunt modulate, în schimb, de neocortexul nostru, care domină creierul omenesc atât ca masă, cât şi ca activitate.

Animalele au trăit şi au supraviețuit fără neocortex şi toate animalele care nu sunt mamifere continuă să o facă în zilele noastre. Putem considera neocortexul omului ca fiind un mare sublimator – astfel, motivația noastră primitivă pentru evitarea unui animal mare de pradă poate fi transformată de neocortexul de azi în îndeplinirea unei sarcini pentru impresionarea şefului; marea vânătoare poate să devină scrierea unei cărți, să spunem despre minte; iar urmărirea reproducerii poate să devină obținerea recunoaşterii publice sau decorarea apartamentului. (Ei bine, această din urmă motivație nu este totdeauna chiar atât de ascunsă.)

Neocortexul este bun, de asemenea, la a ne ajuta să rezolvăm probleme, deoarece poate crea modele precise ale lumii, reflectându-i adevărata natură ierarhică. Însă cel care ne pune în fața tuturor acestor probleme este creierul vechi. Desigur, ca orice birocrație inteligentă, neocortexul rezolvă adeseori problemele care-i sunt alocate prin redefinirea lor. Pe această notă, să trecem în revistă procesarea informației la nivelul creierului vechi.

Calea senzorială

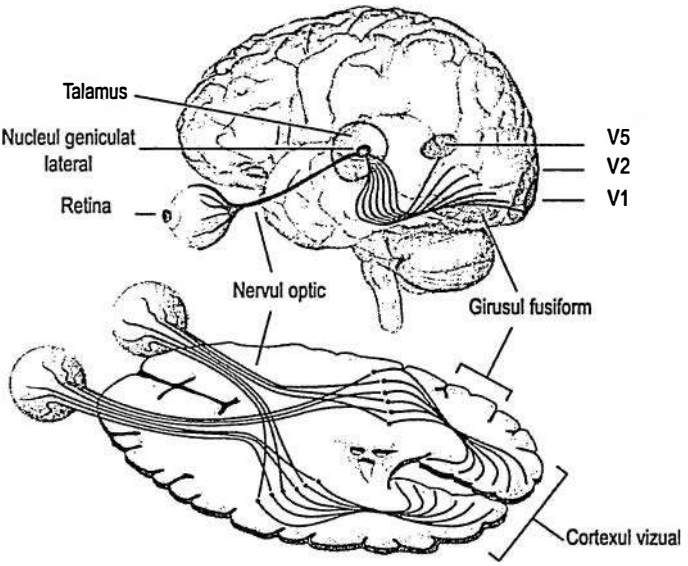
Tablourile, propagate prin mişcare de-a lungul fibrelor nervului optic la creier, sunt cauza vederii.

— Isaac Newton

Fiecare dintre noi trăieşte înăuntrul universului – închisoarea – propriului său creier. Din el proiectează milioane de fibre nervoase senzoriale fragile, în grupuri adaptate unic încercării stărilor energetice ale lumii din jurul nostru: căldura, lumina, forța şi compoziția chimică. Asta e tot ce ştim despre el în mod direct; în rest, totul este deducție logică.

— Vernon Mountcastle[1]

Cu toate că trăim experiența iluziei că primim imagini de înaltă rezoluție de la ochi, ceea ce transmite de fapt la creier nervul optic este doar o serie de schițe şi indicii cu privire la punctele de interes din câmpul nostru vizual. Pe urmă, în esență, visăm lumea din amintirile corticale care interpretează

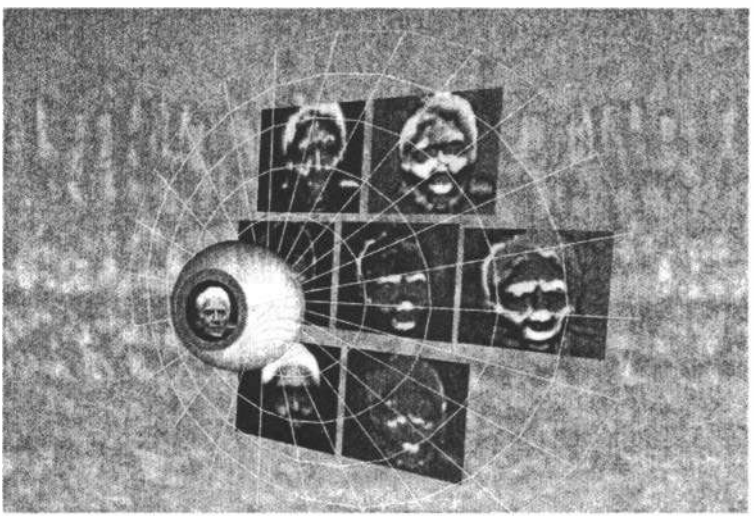


Căile vizuale la nivelul creierului

o serie de filme cu rate ale datelor foarte reduse, care sosesc pe canale paralele. Într-un studiu publicat în revista *Nature*, Frank S. Werblin, profesor de biologie moleculară şi celulară la Universitatea California din Berkeley, şi Boton Roska, doctor în medicină, au arătat că nervul optic conţine între zece şi douăsprezece canale de output, fiecare dintre ele transportând doar un mic volum de informaţie despre o scenă dată.[2] Un grup din ceea ce se numeşte celule ganglionare trimite informaţii referitoare numai la muchii (modificări ale contrastului). Un alt grup detectează doar ariile întinse de culoare uniformă, în vreme ce un al treilea grup este sensibil numai la fundalurile din spatele obiectelor de interes.

„Chiar dacă credem că vedem lumea cât se poate de complet, ceea ce recepţionăm în realitate sunt numai indicii, muchii în spaţiu şi timp", spune Werblin. „Aceste 12 imagini ale lumii reprezintă toată informaţia pe care o vom deţine vreodată referitor la ceea ce există, iar din aceste 12 imagini, care sunt atât de sporadice, noi reconstruim bogăţia lumii vizuale. Sunt curios cum a putut natura să aleagă aceste 12 imagini simple şi cum este posibil ca ele să fie suficiente pentru a ne asigura întreaga informaţie de care s-ar părea că avem nevoie."

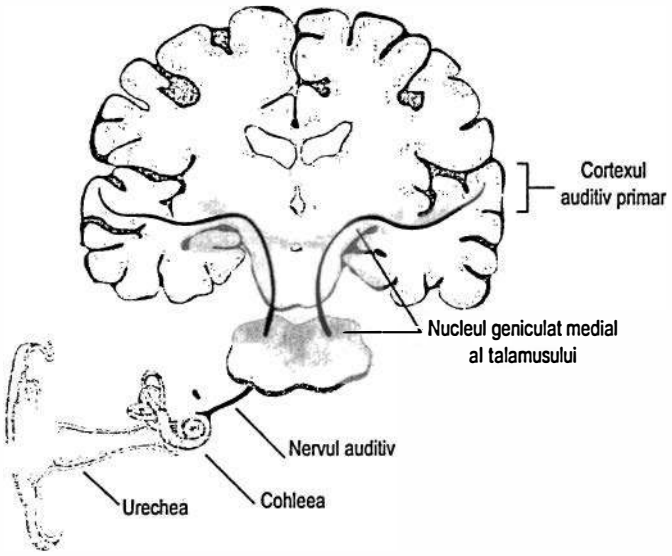
Această reducere a datelor reprezintă ceea ce noi, în domeniul inteligenţei artificiale, numim „codificare sporadică". În crearea sistemelor



Şapte dintre cele douăsprezece „filme" cu conţinut redus de date transmise de nervul optic la creier

artificiale, am descoperit că eliminarea unei mari părți din informațiile de input și reținerea doar a celor mai sugestive detalii asigură rezultate superioare. Altminteri, abilitatea limitată de procesare a informațiilor la nivelul neocortexului (biologică sau de altă natură) ajunge să fie depășită.

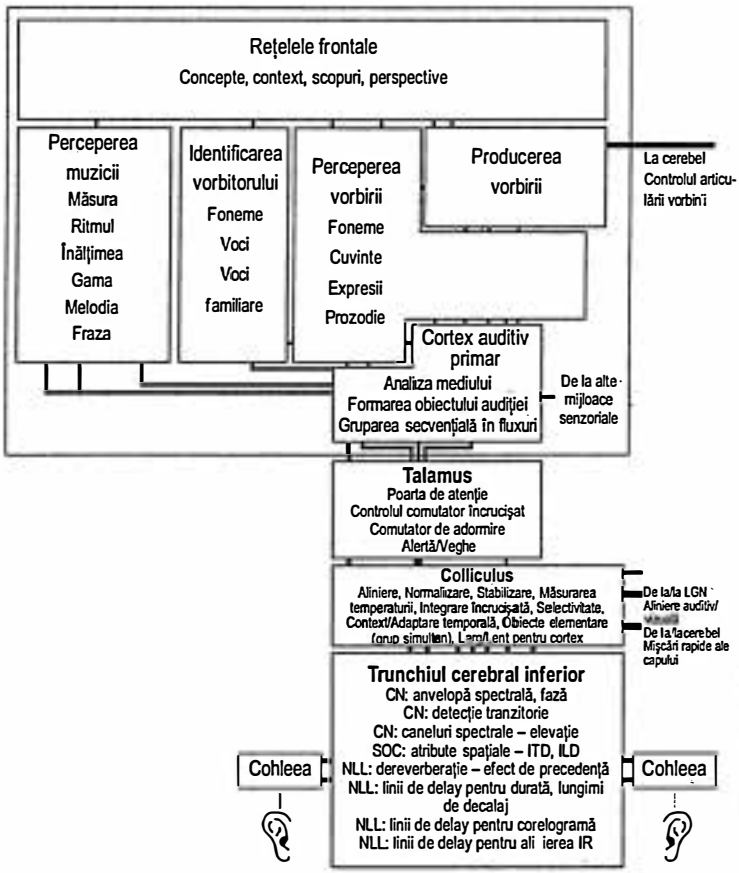
Lloyd Watts, împreună cu echipa lui de cercetare de la Audience, Inc, a creat un model minuțios al procesării informației auditive de către cohleea umană prin intermediul regiunilor subcorticale și apoi prin etajele inferioare ale neocortexului.[3] Această echipă a pus la punct o tehnologie de cercetare care extrage 600 de benzi de frecvență diferite (60 la o octavă) din sunet. Această valoare se apropie mai mult de estimarea de 3.000 de benzi extrase de cohleea umană (în comparație cu sistemele comerciale de recunoaștere a vorbirii, care folosesc numai 16 până la 32 de benzi). Utilizând două microfoane și modelul detaliat (plus rezoluția spectrală înaltă) de procesare auditivă, Audience a creat o tehnologie comercială (cu o rezoluție spectrală oarecum mai redusă decât sistemul de cercetare) care elimină eficient zgomotul de fundal din conversații. Această tehnologie este folosită actualmente de multe telefoane celulare populare și e un impresionant exemplu



Calea auditivă la nivelul creierului

de produs comercial bazat pe înțelegerea modului în care sistemul auditiv al omului este capabil să se concentreze asupra unei surse sonore de interes.

Input-urile de la corp (estimate la sute de megabiți pe secundă), inclusiv de la nervii din piele, mușchi, organe și din alte arii, ajung la nivelul superior al măduvei spinării. Aceste mesaje implică mai mult decât comunicarea referitoare la atingere; în plus, transportă informații referitoare la temperatură,



Un model simplificat al procesării auditive atât la nivelul ariilor subcorticale (arii anterioare neocortexului), cât și la nivelul neocortexului, creat de Audience Inc. Imagine adaptată după L. Watts, *Reverse-Engineering the Human Auditory Pathway*, în J. Liu și colab. (sub redacția), WCCI 2012 (Berlin: Springer-Verlag, 2012), pag. 49

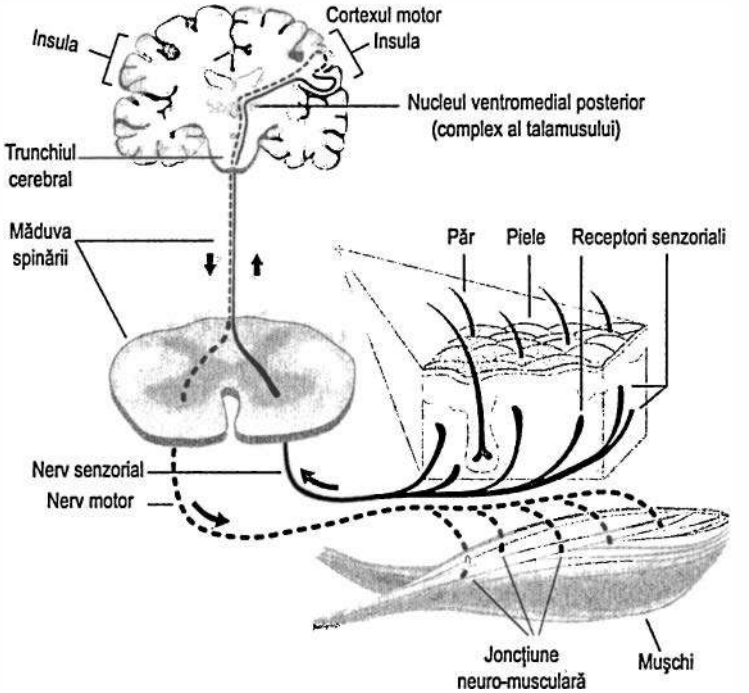
nivelul acizilor (de exemplu, acidul lactic din mușchi), deplasarea alimentelor prin tractul gastrointestinal și multe alte semnale. Aceste date sunt procesate prin trunchiul cerebral și creierul mijlociu. Celule-cheie numite neuroni lamina 1 creează o hartă a corpului, reprezentând starea sa de moment, nu departe de monitoarele folosite de controlorii de trafic pentru urmărirea avioanelor. De aici, datele senzoriale se îndreaptă către o regiune misterioasă numită talamus, care ne aduce la următorul nostru subiect.

Talamusul

> Toată lumea știe ce înseamnă atenția. Înseamnă preluarea de către minte, în formă limpede și vie, a unuia din mai multe subiecte simultan posibile sau șiruri de gânduri. Focalizarea, concentrarea, conștientizarea țin de esența ei. Presupune retragerea de la unele obiecte pentru aprecierea eficientă a altora.
>
> **— William James**

De la creierul mijlociu, informația senzorială este transmisă printr-o regiune de mărimea unei nuci, numită nucleu ventromedial posterior (VMpo) al talamusului, care computează reacțiile complexe la stări ale corpului, cum ar fi „E un gust îngrozitor", „Ce duhoare" sau „Pata aceea de lumină e stimulatoare". Informația procesată din ce în ce mai mult sfârșește în două regiuni ale neocortexului, numite insula. Aceste structuri, de dimensiunile unor degete mici, sunt localizate în partea stângă și în cea dreaptă a neocortexului. Doctorul Arthur Craig de la Barrow Neurological Institute din Phoenix numește VMpo și cele două regiuni ale insulei „sistem care reprezintă eul material".[4]

Printre celelalte funcții ale sale, talamusul este considerat poarta prin care informațiile preprocesate ajung în neocortex. În plus față de informația tactilă care trece prin nucleul ventromedial posterior, informația procesată de la nervul optic (care, după cum am subliniat anterior, a fost deja substanțial transformată) este transmisă într-o regiune a talamusului numită nucleu geniculat lateral, care o trimite apoi la regiunea V1 a neocortexului. Informația de la receptorul auditiv trece prin nucleul geniculat medial al talamusului, îndreptându-se către regiunile auditive prime ale neocortexului. Toate datele



Regiunea căilor senzorial-tactile la nivelul creierului

noastre senzoriale (cu excepția, s-ar părea, a sistemului olfactiv, care folosește bulbii olfactivi) trece prin regiuni specifice ale talamusului.

Cel mai important rol al talamusului, însă, este comunicarea continuă cu neocortexul. Recognizatorii de modele de la nivelul neocortexului trimit rezultate provizorii la talamus și primesc răspunsuri în principal folosind semnale reciproce atât excitatoare, cât și inhibitoare, de la stratul VI al fiecărui recognizator. Reține, cititorule, că nu este vorba de mesaje fără fir, așadar trebuie să existe un volum extraordinar de legături materiale (sub formă de axoni) între toate regiunile neocortexului și talamus. Gândește-te la volumul de spațiu (în termeni de masă fizică de conexiuni necesare) pentru sutele de milioane de recognizatori de modele de la nivelul neocortexului, care trebuie să se consulte permanent cu talamusul.[5]

Așadar, despre ce vorbesc sutele de milioane de recognizatori de modele ai neocortexului cu talamusul? Aparent, este o conversație

importantă, deoarece leziunile profunde bilaterale ale principalelor regiuni ale talamusului pot să conducă la o stare prelungită de pierdere a cunoștinței. O persoană care prezintă leziuni ale talamusului mai poate să aibă activitate la nivelul neocortexului, în sensul că gândirea automată prin asociere mai poate încă funcționa. Dar gândirea direcționată – genul care ne face să ne ducem la culcare, să urcăm în mașină și să ne așezăm la birou la slujbă – nu funcționează în absența talamusului. Într-un caz celebru, Karen Ann Quinlan, în vârstă de douăzeci și unu de ani, a suferit un infarct miocardic și manifesta insuficiență respiratorie, așa că a rămas într-o stare apatică, aparent vegetativă, timp de zece ani. După moarte, autopsia a dezvăluit faptul că neocortexul ei era normal, dar talamusul fusese distrus.

Pentru a-și îndeplini rolul în privința abilității noastre de a ne direcționa atenția, talamusul se bazează pe cunoașterea structurată conținută de neocortex. El poate parcurge o listă (stocată în neocortex), permițându-ne să urmărim un șir de gânduri sau să respectăm un plan de acțiune. Aparent, suntem capabili să avem, în același timp, în jur de patru subiecte active în memoria de lucru, câte două în fiecare emisferă, potrivit recentelor cercetări ale specialiștilor în neuroștiință de la MIT Picower Institute for Learning and Memory.[6] Problema dacă talamusul este responsabil de neocortex sau invers e departe de a fi clarificată, însă nu suntem în măsură să funcționăm în lipsa ambelor.

Hipocampul

Fiecare emisferă cerebrală are un hipocamp, o mică regiune care seamănă cu un căluț de mare, ghemuit medial în lobul temporal. Principala sa funcție este să-și amintească evenimentele noi. De vreme ce informația senzorială trece prin neocortex, de el depinde să stabilească dacă o experiență este nouă pentru a o prezenta hipocampului. Și face acest lucru fie nerecunoscând un anumit set de caracteristici (de exemplu, un chip nou), fie realizând faptul că o situație altminteri familiară are într-un moment actual atribute unice (cum ar fi acela că soția ta are o mustață falsă).

Hipocampul este capabil să rețină asemenea situații, cu toate că s-ar părea că face acest lucru în primul rând prin anumite indicatoare din neocortex. Prin urmare, amintirile din hipocamp sunt stocate și ca modele de nivel inferior, care au fost anterior recunoscute și depozitate în neocortex.

Pentru ca animalele care nu au neocortex să poată modula experiențele senzoriale, hipocampul pur şi simplu memorează informația de la simțuri, deşi aceasta e trecută deja prin preprocesarea senzorială (de exemplu, transformarea efectuată de nervul optic). Deşi hipocampul se foloseşte de neocortex (dacă creierul respectiv beneficiază de aşa ceva) ca de un carnețel de notițe, memoria sa (de indicatoare din neocortex) nu este în mod inerent ierarhică. Animalele lipsite de neocortex pot, prin urmare, să memoreze diverse lucruri folosindu-şi hipocampul, dar amintirile lor nu sunt ierarhice.

Capacitatea hipocampului este limitată, aşadar memorarea efectuată de el este de scurtă durată. Hipocampul transferă o anumită succesiune de modele din memoria sa de scurtă durată în memoria ierarhică de lungă durată a neocortexului, repetându-i acestuia, iar şi iar, respectiva succesiune de memorie. Prin urmare, avem nevoie de hipocamp pentru a memora lucruri şi abilități noi (deşi abilitățile strict motorii par să folosească un mecanism diferit). O persoană care a suferit tulburări la ambele còpii ale hipocampului îşi păstrează amintirile existente, dar nu-şi mai poate crea altele noi.

Theodore Berger, specialist în neuroştiință la Universitatea South California, împreună cu colaboratorii săi, a realizat un model al hipocampului de şobolan şi a experimentat cu succes implantarea unuia nou. În cadrul unui studiu ale cărui rezultate au fost comunicate în anul 2011, oamenii de ştiință de la Universitatea South California au blocat la şobolani, cu ajutorul medicamentelor, anumite comportamente învățate. Folosindu-se de un hipocamp artificial, şobolanii au fost repede capabili să reînvețe comportamentul. „Închizi comutatorul şi şobolanul îşi aminteşte. Îl deschizi şi şobolanul uită", scria Berger referindu-se la posibilitatea lui de a controla de la distanță implantul neuronal. În cadrul unui alt experiment, oamenii de ştiință au permis hipocampului lor artificial să funcționeze în paralel cu cel natural al şobolanului. Rezultatul a fost acela că s-a consolidat abilitatea şobolanilor de a învăța noi abilități. „Aceste studii integrate de modelare experimentală arată, pentru prima oară", explica Berger, „că... o proteză neuronală capabilă de identificarea şi manipularea în timp real a procesului de codificare poate să refacă şi chiar să stimuleze procesele cognitive mnemonice".[7] Hipocampul este una dintre primele regiuni afectate de boala Alzheimer, deci unul dintre scopurile acestor cercetări îl reprezintă punerea la punct a unui implant neuronal pentru oameni, implant care va rezolva această primă fază a leziunilor provocate de boală.

Cerebelul

Există două modalităţi la care poţi apela pentru a încerca să prinzi o minge în zbor. Poţi să rezolvi ecuaţiile diferenţiale simultane complexe care controlează mişcarea mingii, dar şi alte ecuaţii care-ţi guvernează unghiul particular sub care vezi mingea, iar apoi să computezi şi mai multe ecuaţii pentru mişcarea corpului, braţului şi mâinii pentru a te afla în locul potrivit la momentul potrivit.

Însă nu aceasta este abordarea pe care o adoptă creierul tău. În esenţă, el simplifică problema, reducând o mulţime de ecuaţii la un simplu model al orientării, apreciind locul în care pare să se afle mingea în câmpul vizual şi rapiditatea cu care se deplasează aceasta. Creierul face acelaşi lucru şi cu mâna ta, realizând predicţii liniare ale poziţiei aparente a mingii în câmpul vizual şi ale mişcărilor mâinii. Scopul, desigur, este acela ca el să se asigure că cele două elemente se întâlnesc în acelaşi punct din spaţiu şi din timp. Dacă mingea pare să cadă prea repede şi mâna pare să se mişte prea încet, creierul îţi face mâna să se mişte mai repede, astfel încât traiectoriile să coincidă. Această soluţie gen „nodul gordian" pentru ceea ce altminteri ar reprezenta o problemă insolvabilă matematic se numeşte funcţie de bază şi e îndeplinită de cerebel, o regiune cu forma unui bob de fasole şi de mărimea unei mingi de baseball, situată deasupra trunchiului cerebral.[8]

Cerebelul este o regiune a creierului vechi care a controlat cândva practic toate mişcările hominizilor. El încă mai conţine jumătate dintre neuronii creierului, deşi cei mai mulţi sunt relativ mici, astfel încât regiunea reprezintă numai în jur de 10 la sută din greutatea creierului. Pe de altă parte, cerebelul este o dovadă de masivă repetiţie în construcţia creierului. Există puţine informaţii în genom cu privire la construcţia lui, structura sa fiind un model de mai mulţi neuroni care se repetă de miliarde de ori. La fel ca în cazul neocortexului, el prezintă o uniformitate a întregii sale structuri.[9]

Cea mai mare parte a funcţiilor de control asupra activităţii musculare a fost preluată de neocortex, folosindu-se aceiaşi algoritmi de recunoaştere a modelelor care sunt utilizaţi şi pentru percepţie şi cogniţie. În cazul mişcării, la funcţia neocortexului ne putem referi mai adecvat prin sintagma „implementarea modelului". Într-adevăr, neocortexul se foloseşte de memoria din cerebel pentru înregistrarea scenariilor delicate ale mişcărilor – de exemplu, semnătura şi anumite înfloriturii de expresie artistică, bunăoară în muzică şi

dans. Studiile asupra rolului cerebelului în timpul învățării scrierii de mână de către copii dezvăluie faptul că celulele Purkinje de la nivelul cerebelului iau de fapt mostre de mişcări, fiecare celulă fiind sensibilă la o anumită mostră.[10] Deoarece mişcările noastre, în cea mai mare parte a lor, sunt controlate de neocortex, mulți oameni se pot descurca bine în condițiile unor dizabilități evidente relativ modeste, ba chiar şi în cazul lezării cerebelului, doar că mişcările pot deveni mai puțin grațioase.

Neocortexul mai poate recurge la cerebel şi pentru a-i folosi abilitatea de computare a funcțiilor de bază în timp real, în vederea anticipării rezultatelor acțiunilor pe care le avem în vedere, dar pe care nu le-am realizat încă (şi e posibil să nici nu le efectuăm vreodată), precum şi a acțiunilor sau a posibilelor acțiuni ale celorlalți. Este încă o dovadă a existenței predictorilor liniari înnăscuți incluşi în creier.

S-au înregistrat progrese substanțiale în simularea cerebelului sub aspectul capacității sale de a răspunde dinamic la indicii senzoriale, folosind funcțiile de bază pe care le-am descris anterior atât în simulările inversate (bazate pe modele biochimice), cât şi în simulările directe (bazate pe modele matematice ale operării la nivelul cerebelului a fiecărei unități repetate).[11]

Plăcere şi frică

> Frica este principala sursă de superstiții şi una dintre principalele surse de cruzime. Înfrângerea fricii este începutul înțelepciunii.
>
> **— Bertrand Russell**

> Simte frica şi alung-o.
>
> **— Susan Jeffers**

Dacă neocortexul este bun la rezolvarea problemelor, atunci care-i principala problemă pe care vrem să o rezolvăm? Problema pe care evoluția a încercat întotdeauna s-o soluționeze o reprezintă supraviețuirea speciilor. Aceasta se traduce prin supraviețuirea individului şi fiecare dintre noi îşi foloseşte propriul neocortex ca să interpreteze acest lucru într-o multitudine de moduri. Pentru a supraviețui, animalele sunt nevoite să-şi procure următoarea masă şi, concomitent, să evite să devină masa altui animal. De asemenea, trebuie să se reproducă. Creierul timpuriu şi-a dezvoltat sisteme

de frică și plăcere care să recompenseze îndeplinirea acestor nevoi fundamentale, laolaltă cu comportamente care le facilitau. Pe măsură ce mediul și speciile concurente s-au modificat, evoluția biologică a făcut schimbările corespunzătoare. Odată cu apariția gândirii ierarhizate, satisfacția pornirilor fundamentale a devenit mai complexă, fiind de-acum supusă vastului complex de idei din idei. Însă, în ciuda considerabilei modulări de către neocortex, creierul vechi este în continuare viu și încă ne mai motivează cu plăcerea și cu frica.

O regiune asociată cu plăcerea este nucleus accumbens. În celebrele experimente realizate în anii 1950, șobolanii s-au arătat capabili să își stimuleze direct această mică regiune (prin apăsarea unei pedale care activa electrozii implantați), preferând această activitate oricărei alteia, inclusiv raportului sexual sau hrănirii, iar în cele din urmă epuizându-se și înfometându-se până la moarte.[12] La om, mai există și alte regiuni implicate în plăcere, cum ar fi nucleus pallidum ventral și, desigur, neocortexul însuși.

Plăcerea este provocată și de compuși chimici cum ar fi dopamina și serotonina. Nu intră în scopurile acestei cărți discutarea detaliată a acestor sisteme, dar e important să recunoaștem că am moștenit mecanismele cu pricina de la verișoarele noastre premamifere. Menirea neocortexului este aceea de a ne permite să fim stăpânii plăcerii și fricii, nu sclavii lor. Cum însă suntem adeseori supuși unor comportamente de dependență, neocortexul nu are întotdeauna succes în acest sens. Dopamina, mai ales, este un neurotransmițător implicat în trăirea plăcerii. Când ni se întâmplă ceva plăcut – câștigăm la loterie, obținem recunoașterea colegilor, primim o îmbrățișare de la o persoană iubită sau chiar înregistrăm realizări subtile, cum ar fi să-l facem pe un prieten să râdă la o glumă –, trăim experiența unei descărcări de dopamină. Uneori și noi, asemenea șobolanilor care au murit stimulându-și excesiv nucleus accumbens, folosim o scurtătură pentru obținerea acestor accese de plăcere, ceea ce nu e întotdeauna o idee bună.

Jocurile de noroc, de exemplu, pot descărca dopamină cel puțin atunci când câștigi, dar asta depinde de inerenta lipsă de predictibilitate a situației. Jocurile de noroc pot sluji scopul descărcării dopaminei pentru un timp, dar, dat fiind faptul că șansele sunt intenționat create pentru a lucra împotriva ta (altminteri nu ar funcționa modelul de afacere al unui cazinou), pot deveni ruinătoare ca strategie regulată. Pericole similare sunt asociate cu oricare alt comportament de dependență. O anumită mutație a genei receptorului de dopamină D2 provoacă sentimente de plăcere deosebit de puternice la

primele experienţe cu substanţe adictive şi cauzează comportamente ce dau dependenţă, dar, după cum este bine ştiut (însă nu întotdeauna ţinut minte), capacitatea acestor substanţe de a produce plăcere la utilizările ulterioare de obicei se reduce. O altă mutaţie genetică rezidă în faptul că unele persoane nu beneficiază de descărcarea unor cantităţi normale de dopamină în urma realizărilor cotidiene, ceea ce poate de asemenea să conducă la căutarea unor experienţe stimulatoare prin activităţi care dau dependenţă. Minoritatea celor care manifestă aceste tendinţe genetice de dependenţă creează o enormă problemă socială şi medicală. Chiar şi persoanele care reuşesc să evite comportamentele sever dependente se luptă cu echilibrarea recompenselor la descărcările de dopamină.

Serotonina este un neurotransmiţător care joacă un rol major în reglarea stărilor de spirit. La niveluri mai înalte, este asociată cu sentimente de bine şi mulţumire. Serotonina are şi alte funcţii, inclusiv modularea tăriei sinaptice, a apetitului, somnului, dorinţei sexuale şi digestiei. Medicamentele antidepresive, cum ar fi inhibitorii selectivi de refacere a serotoninei (care tind să sporească nivelul de serotonină disponibilă pentru receptori), tind să aibă efecte cu bătaie lungă – nu toate dezirabile (cum ar fi suprimarea libidoului). Spre deosebire de acţiunile de la nivelul neocortexului, unde recunoaşterea modelelor şi activarea axonilor afectează concomitent doar un număr redus de circuite neocorticale, antidepresivele afectează regiuni întinse ale creierului sau chiar întregul sistem nervos.

Fiecare emisferă a creierului uman are o amigdală, care este o regiune în formă de migdală, conţinând mai mulţi lobi mici. Amigdala e şi ea parte a creierului vechi şi e implicată în procesarea mai multor tipuri de răspunsuri emoţionale, cel mai notabil dintre ele fiind frica. La animalele premamifere, anumiţi stimuli preprogramaţi care semnalează pericolul se transmit direct amigdalei, care, la rândul ei, declanşează mecanismul „Luptă sau fugi". La om, actualmente, amigdala este importantă pentru transmiterea percepţiei de pericol la neocortex. Un comentariu negativ al şefului, de exemplu, poate declanşa un asemenea răspuns prin generarea fricii de pierdere a slujbei (sau poate nu, dacă ai încredere într-un plan B). Din momentul în care amigdala decide că te pândeşte un pericol, survine o succesiune ancestrală de evenimente. Amigdala îi transmite glandei pituitare să descarce un hormon denumit ACTH (adrenocorticotropină). Acesta, la rândul său, stimulează descărcarea în sânge a hormonului de stres cortizol, localizat în glandele suprarenale, iar prezenţa acestuia se concretizează în asigurarea unei energii mai

mari la nivelul sistemului muscular şi nervos. Glandele suprarenale secretă, de asemenea, adrenalină şi noradrenalină, care inhibă sistemele digestiv, imunitar şi reproducător (imaginându-şi că nu sunt procese de mare prioritate într-o situație de urgență). Cresc în schimb nivelurile tensiunii arteriale, glicemiei, colesterolului şi fibrinogenului (care accelerează coagularea sângelui). Bătăile inimii şi respirația se întețesc. Chiar şi pupilele se dilată, astfel încât să existe o mai bună acuitate vizuală pentru observarea inamicului sau pentru identificarea căii de scăpare. Toate acestea sunt foarte utile dacă întâlneşti pe neaşteptate un pericol, cum ar fi un prădător. Este foarte bine cunoscut faptul că, în lumea zilelor noastre, activarea cronică a mecanismului „Luptă sau fugi" poate conduce la afecțiuni permanente sub forma hipertensiunii, hipercolesterolemiei ş.a.

Sistemul global al nivelurilor neurotransmițătorilor cum ar fi serotonina şi al hormonilor cum ar fi dopamina este foarte complicat şi am putea să ocupăm tot restul cărții de față cu acest subiect (cum s-a şi întâmplat cu foarte multe cărți), dar merită să subliniez că lățimea de bandă a informației (rata de procesare a informației) din acest sistem e foarte redusă comparativ cu lățimea de bandă a neocortexului. Există un număr limitat de substanțe implicate, iar nivelurile acestor compuşi chimici tind să se modifice încet şi sunt relativ universale în tot creierul, spre deosebire de neocortex, care e format din sute de miliarde de celule ce se pot schimba rapid.

Este corect să spun că experiențele noastre emoționale se produc atât la nivelul creierului vechi, cât şi la cel al creierului nou. Gândirea are loc în creierul nou (neocortexul), dar sentimentele apar în ambele. Orice imitare a comportamentului uman trebuie, prin urmare, să conțină modele ale ambelor. Pe de altă parte, dacă urmărim numai inteligența cognitivă a omului, neocortexul este suficient. Putem înlocui creierul vechi cu o motivație mai directă a unui neocortex nonbiologic pentru obținerea scopurilor pe care i le atribuim. De exemplu, în cazul lui Watson, scopul a fost formulat simplu: Dă răspunsurile corecte la întrebările de la *Jeopardy!* (cu toate că acestea erau modulate şi mai mult de un program care înțelegea cum se joacă *Jeopardy!).* În cazul noului sistem Watson, care este pus la punct, în colaborare, de Nuance şi IBM pentru achiziția de cunoştințe medicale, scopul este ca acesta să ajute la tratarea afecțiunilor omului. Sistemele viitoare vor putea să aibă scopuri cum ar fi vindecarea propriu-zisă a bolilor şi eliminarea sărăciei. În mare măsură, lupta plăcere-frică este deja străveche pentru om, creierul vechi evoluând cu mult înainte chiar şi de apariția

societății umane primitive; într-adevăr, în marea sa parte, acest creier este reptilian.

În creierul omului se dă o luptă continuă: aparține controlul creierului vechi sau creierului nou? Creierul vechi încearcă să facă agenda cu controlul experiențelor de plăcere şi frică, în vreme ce creierul nou se străduieşte din răsputeri să înțeleagă algoritmii relativ primitivi ai creierului vechi şi să îi manipuleze după propriile interese. Reține că amigdala este incapabilă să evalueze singură pericolul – creierul omului se bazează pe neocortex pentru a judeca un asemenea lucru. Persoana aceea e prieten sau duşman, o persoană iubită sau o amenințare? Numai neocortexul poate să decidă.

În măsura în care nu suntem implicați direct într-o luptă pe viață şi pe moarte şi nu vânăm ca să ne hrănim, am reuşit să ne sublimăm cel puțin parțial pornirile ancestrale în sensul unor acțiuni mai temerare. Pe această notă, vom discuta creativitatea şi iubirea în capitolul care urmează.

Capitolul 6

ABILITĂȚILE TRANSCENDENTE

Aceasta este religia mea simplă. Nu e nevoie de temple; nu e nevoie de o filosofie complicată. Propriul creier şi propria inimă ne sunt templul; filosofia este generozitatea.

— Dalai Lama

Mâna mi se mişcă pentru că anumite forțe – electrice, magnetice sau orice altă „forță nervoasă" – îmi sunt imprimate în creier. Forței nervoase depozitate în creier probabil că i s-ar putea da de urmă, dacă ştiința ar fi completă, în forțele chimice puse la dispoziția creierului de către sânge şi e obținută, în cele din urmă, din alimentele pe care le mănânc şi din aerul pe care îl respir.

— Lewis Carroll

Tot în neocortex îşi au sediul şi gândurile noastre emoționale, dar sunt influențate de părți ale creierului care merg de la regiuni ale creierului vechi, precum amigdala, până la structuri ale creierului recente din punct de vedere evolutiv, cum ar fi neuronii fusiformi, care par să joace un rol-cheie în emoțiile de ordin superior. Spre deosebire de structurile recursive regulate şi logice care se găsesc în cortexul cerebral, neuronii fusiformi prezintă contururi şi conexiuni foarte neregulate. Sunt cei mai mari neuroni din creierul uman, acoperind toată întinderea lui, şi sunt profund interconectați, cu sute de mii de legături care unesc unele de altele diferite porțiuni ale neocortexului.

După cum am menționat anterior, insula ajută la procesarea semnalelor senzoriale, dar joacă un rol-cheie şi în emoțiile de ordin superior. Aceasta este regiunea în care îşi au originea neuronii fusiformi. Scanările pe bază de rezonanță magnetică în condiții funcționale (RMN) au dezvăluit faptul că aceste celule sunt deosebit de active atunci când persoana se confruntă cu sentimente cum ar fi iubirea, furia, tristețea şi cu dorința sexuală. Situațiile care le activează intens le includ şi pe cele în care o femeie îşi priveşte partenerul sau îşi aude copilul plângând.

Neuronii fusiformi au filamente neuronale lungi, denumite dendrite apicale, care sunt capabile să se conecteze cu regiuni îndepărtate ale neocortexului. O asemenea interconectare „profundă", prin care anumiți neuroni asigură conectarea la distanțe de mai multe regiuni, reprezintă o caracteristică tot mai evidentă pe măsură ce urcăm pe scara evoluției. Nu este surprinzător faptul că neuronii fusiformi, atât de implicați în emoții și judecăți morale, au acest nivel de conectare, dată fiind capacitatea reacțiilor emoționale de nivel superior de a atinge diverse aspecte și gânduri.

Datorită legăturilor lor cu numeroase regiuni ale creierului, emoțiile de nivel superior pe care le procesează neuronii fusiformi sunt afectate de toate regiunile noastre legate de percepție și cogniție. Este important să subliniez faptul că aceste celule nu realizează o rezolvare rațională a problemelor, motiv pentru care nici nu avem un control rațional asupra răspunsurilor noastre la muzică sau la momentele când ne îndrăgostim. Pe de altă parte, restul creierului e puternic angajat în încercarea de a desluși misterioasele noastre emoții de ordin superior.

Există relativ puțini neuroni fusiformi: doar în jur de 80.000, din care aproximativ 45.000 în emisfera dreaptă și 35.000 în emisfera stângă. Această disparitate explică parțial percepția potrivit căreia inteligența emoțională este domeniul emisferei drepte a creierului, deși disproporția e modestă. Gorilele au în jur de 16.000 de asemenea celule, cimpanzeul pitic are cam 2.100, iar cimpanzeii – în jur de 1.800. De la alte mamifere, ele lipsesc cu desăvârșire.

Antropologii cred că neuronii fusiformi și-au făcut apariția în urmă cu 10 până la 15 milioane de ani, la deocamdată nedescoperitul strămoș al maimuțelor și al omului și că au crescut rapid numeric acum aproximativ 100.000 de ani. Fapt interesant, neuronii fusiformi nu există la nou-născutul uman și încep să apară abia în jurul vârstei de patru luni, iar numărul lor crește semnificativ între vârstele de unu și trei ani. Capacitatea copiilor de a face față temelor morale și de a percepe emoții de ordin superior cum ar fi iubirea se dezvoltă exact în aceeași perioadă.

Aptitudinile

Wolfgang Amadeus Mozart (1756-1791) a scris un menuet pe când avea cinci ani. La vârsta de șase ani, a cântat în fața împărătesei Maria

Tereza, la curtea imperială de la Viena. A compus șase sute de piese, inclusiv patruzeci și una de simfonii, înaintea morții sale survenite la vârsta de treizeci și cinci de ani și este considerat îndeobște cel mai mare compozitor european de muzică clasică. S-ar putea spune că a avut aptitudini pentru muzică.

Așadar, ce înseamnă asta în contextul Teoriei minții bazate pe recunoașterea modelelor? În mod evident, ceea ce considerăm că reprezintă o aptitudine este produsul cultivării – cu alte cuvinte, al influențelor mediului și al altor oameni. Mozart s-a născut într-o familie muzicală. Tatăl lui, Leopold, era compozitor și *kapellmeister* (conducător muzical, literal) al orchestrei curții arhiepiscopului de Salzburg. Tânărul Mozart a fost scufundat în muzică și tatăl lui a început să-l învețe vioara și clavirul (un instrument cu claviaturi) de la vârsta de trei ani.

Pe de altă parte însă, numai influențele mediului nu pot să explice complet geniul lui Mozart. Există aici, în mod evident, și o componentă naturală. Ce formă îmbracă aceasta? După cum am scris în Capitolul 4, diferite regiuni ale neocortexului au ajuns să fie optimizate (prin evoluția biologică) pentru anumite tipuri de modele. Cu toate că algoritmul de bază al recunoașterii modelelor este uniform în tot neocortexul, dat fiind faptul că anumite tipuri de modele tind să acopere anumite regiuni (chipurile la nivelul girusului fusiform, de exemplu), aceste regiuni ajung să fie mai performante la procesarea modelelor asociate.

În orice caz, există numeroși parametri care guvernează modul cum este aplicat propriu-zis algoritmul la nivelul fiecărui modul. De exemplu, cât de mare trebuie să fie asemănarea pentru ca modelul să fie recunoscut? Cum se modifică acest prag dacă un modul de nivel superior trimite un semnal potrivit căruia este „așteptat" modelul său? Cum sunt apreciați parametrii de mărime? Acești factori și alții au fost stabiliți diferit în regiuni diferite, pentru a reprezenta un avantaj pentru anumite tipuri de modele. În cercetările noastre pe metode similare în domeniul inteligenței artificiale, am observat același fenomen și am folosit simulări ale evoluției pentru optimizarea acestor parametri.

Dacă anumite regiuni pot fi optimizate pentru diferite tipuri de modele, rezultă că creierele individuale variază și ele, prin capacitatea de învățare, de recunoaștere și de creare a anumitor tipuri de modele. De exemplu, un creier poate avea aptitudini înnăscute pentru muzică, fiind capabil într-o măsură mai mare să recunoască modelele ritmice, ori poate fi mai înclinat

să înțeleagă aranjamentele geometrice ale armoniilor. Fenomenul înălțimii perfecte (capacitatea de a recunoaşte şi a reproduce o înălțime fără un reper exterior), corelat cu talentul muzical, pare să aibă o bază genetică, deşi capacitatea aceasta trebuie cultivată, ceea ce înseamnă că este foarte posibil să fie o combinație de natură şi cultură. Baza genetică a înălțimii perfecte e posibil să rezide în afara neocortexului, în preprocesarea informației auditive, pe când aspectul învățat rezidă în neocortex.

Există şi alte abilități care contribuie la gradul de competență, indiferent dacă este vorba de genul rutinier sau de geniul legendar. Abilitățile neocorticale – de exemplu, capacitatea neocortexului de a controla semnalele de frică pe care le generează amigdala (când se confruntă cu dezaprobarea) – joacă un rol semnificativ, la fel ca alte atribute, cum ar fi încrederea, abilitățile organizaționale şi capacitatea de a-i influența pe alții. O abilitate foarte importantă pe care am scos-o în evidență mai devreme este curajul de a urmări idei care merg împotriva preceptelor ortodoxe. Invariabil, oamenii pe care îi considerăm genii şi-au luat în considerare propriile experimente mentale în moduri care inițial nu au fost înțelese sau apreciate de către confrații lor. Deşi Mozart şi-a câştigat recunoaşterea în cursul vieții, cea mai mare parte a adulației a venit abia mai târziu. A murit sărac, a fost înmormântat într-o groapă comună, iar la funeraliile lui au participat doar alți doi muzicieni.

Creativitatea

> Creativitatea este un drog fără care nu pot să trăiesc.
>
> **— Cecil B. DeMille**

> Problema nu este niciodată cum să-ți bagi în minte idei noi, inovatoare, ci cum să le scoți pe cele vechi. Fiecare minte este o clădire plină de mobilă veche. Fă-ți curățenie într-un colț al minții şi instantaneu îl va umple creativitatea.
>
> **— Dee Hock**

> Umanitatea poate fi cât se poate de rece față de cei ai căror ochi văd lumea altfel.
>
> **— Eric A. Burns**

Creativitatea poate rezolva aproape orice problemă. Actul creator, înfrângerea obișnuinței de către originalitate, biruiește orice.

— George Lois

Un aspect-cheie al creativității este procesul de găsire a marilor metafore – simboluri care reprezintă altceva. Neocortexul e o mare mașină de produs metafore, fapt ce explică de ce suntem o specie extraordinar de creativă. Oricare dintre cele aproximativ 300 de milioane de recognizatori de modele din neocortexul nostru recunoaște, definește un model și îi dă un nume. Asta, în cazul modulelor neocorticale de recunoaștere a modelelor, înseamnă pur și simplu că un axon iese din recognizatorul de modele și emite un impuls atunci când găsește modelul respectiv. La rândul lui, acest simbol devine parte a altui model. Fiecare dintre aceste modele este, în esență, o metaforă. Recognizatorii pot să emită impulsuri și de 100 de ori pe secundă, așadar avem potențialul de recunoaștere a unui număr de 30 de miliarde de metafore pe secundă. Desigur, nu toate modulele emit semnale la fiecare ciclu – însă este corect să se spună că, într-adevăr, recunoaștem milioane de metafore pe secundă.

Desigur, unele metafore sunt mai semnificative decât altele. Darwin și-a dat seama că intuiția lui Charles Lyell, potrivit căreia modificările extrem de graduale provocate de o picătură de apă pot săpa canioane imense, constituie o metaforă sugestivă a modului în care o picătură de mici modificări evolutive, pe parcursul miilor de generații, poate produce modificările imense în diferențierea speciilor. Experimentele mintale, cum ar fi cel pe care l-a folosit Einstein pentru a pune în lumină adevărata semnificație a experimentului Michelson-Morley, sunt toate metafore, în sensul că reprezintă un „lucru considerat reprezentativ sau simbolic pentru altceva", ca să citez o definiție de dicționar.

Vezi vreo metaforă în *Sonetul 73* de Shakespeare?

Acea vreme-a anului ce-n mine ai vedea-o
Când frunze galbene, ori niciuna, ori câteva, mai stau
Pe cele crengi ce tremură de frig
Strane pustii năruite, în care odinioară păsări cântau.
În mine vezi amurgul unei asemenea zile
Cum dup-apus se pierde-n miazănoapte,
Pe care-ncet-încet noaptea neagră trebuie să-l răpească,

A doua faţ-a morţii ce pecetluieşte totul în odihnă.
În mine vezi sclipirile unui asemenea foc
Ce pe cenuşa tinereţii trebuie să ardă
Ca pe-un pat al morţii pe care trebuie să se sfârşească
Înghiţit de ceea ce a fost hrănit.
Asta vezi şi iubirea ţi-o întăreşte.
Mult să iubeşti ce-aici trebuie să laşi.

În acest sonet, poetul foloseşte metafore extinse pentru a-şi descrie înaintarea în vârstă. Vârsta lui e precum toamna târzie, „când frunze galbene, ori niciuna, ori câteva, mai stau". Vremea e rece şi păsările nu mai pot rămâne pe ramuri, pe care poetul le numeşte „strane pustii năruite". Vârsta lui este asemenea apusului: „cum dup-apus se pierde-n miazănoapte/ pe care-ncet-încet noaptea neagră trebuie să-l răpească". El este restul unui foc „ce pe cenuşa tinereţii trebuie să ardă". Pe bună dreptate, întregul limbaj este în ultimă instanţă o metaforă, cu expresii unele mai memorabile decât altele.

Găsirea unei metafore reprezintă procesul de recunoaştere a unui model în ciuda diferenţelor de detaliu şi context – o activitate pe care o efectuăm în mod curent, în fiecare moment al vieţii noastre. Salturile metaforice pe care le considerăm semnificative tind să aibă loc în spaţiile dintre diferite discipline. Pe de altă parte, acţionând împotriva forţei esenţiale a creativităţii, manifestăm tendinţa atotprezentă spre o tot mai profundă specializare în ştiinţe (şi în aproape toate celelalte domenii, de altfel). După cum a scris matematicianul american Norbert Wiener (1894-1964) în lucrarea sa fundamentală *Cibernetica*, publicată în anul în care m-am născut eu (1948):

> Există domenii de cercetare ştiinţifică, după cum vom vedea în cuprinsul acestei cărţi, care au fost explorate din perspectivele diferite ale matematicilor pure, statisticii, ingineriei electrice şi neurofiziologiei; domenii în care absolut toate noţiunile capătă un nume diferit de la fiecare grup şi în care cercetări foarte importante au fost triplate sau cvadruplate, în timp ce alte cercetări importante sunt întârziate de lipsa dintr-un domeniu a unor rezultate care e posibil să fi devenit deja clasice într-un alt domeniu.

Aceste domenii de frontieră oferă cele mai bogate oportunități cercetătorului pregătit. În același timp, sunt cele mai refractare la tehnicile acceptate de atac în masă și de diviziune a muncii.

O tehnică pe care am folosit-o în propriile mele cercetări pentru combaterea specializării tot mai accentuate a fost să adun experți pentru un proiect (de exemplu, cercetările mele de recunoaștere a vorbirii au implicat oameni de știință, lingviști, psihoacusticieni și experți în recunoașterea modelelor, fără să-i mai menționez pe cercetătorii din domeniul computerului) și să-i încurajez pe toți să predea grupului tehnicile și terminologia care le erau particulare. Pe urmă, ne-am descotorosit de toate aceste terminologii și am conceput-o pe a noastră. Invariabil, am găsit metafore dintr-un domeniu care rezolvă probleme în altul.

Un șoarece care găsește o cale de scăpare atunci când ajunge față în față cu pisica din gospodărie – și care poate face acest lucru chiar și dacă situația este pe undeva diferită de tot ceea ce a trăit anterior – dă dovadă de creativitate. Propria noastră creativitate este cu câteva ordine de mărime mai înaltă decât cea a șoarecelui – și implică mult mai multe niveluri de abstracție – deoarece noi avem un cortex mult mai mare, capabil de niveluri mult mai mari de ierarhizare. Prin urmare, un mod de asigurare a unei creativități mult mai mari constă în asamblarea eficientă a unui neocortex mai mare.

O abordare a expansiunii neocortexului disponibil se poate face prin colaborarea mai multor oameni. Operația se realizează în mod curent prin comunicarea dintre oamenii reuniți într-o comunitate pentru rezolvarea unei probleme. Recent, s-au întreprins eforturi pentru folosirea instrumentelor de colaborare online în scopul exploatării forței colaborării în timp real, iar strategia a avut succes în matematică și în alte domenii.[1]

Desigur, următorul pas va consta în extinderea neocortexului însuși, prin echivalentul lui nonbiologic. Acesta va fi actul nostru suprem de creativitate: să creăm capacitatea de a fi creativi. Un neocortex nonbiologic ar fi, în primul și în primul rând, mai rapid și ar putea să caute repede genurile de metafore care i-au inspirat pe Darwin și Einstein. Ar putea să exploreze sistematic toate zonele de suprapunere a frontierelor dintre domeniile noastre de cunoaștere în continuă expansiune.

Unii își exprimă preocuparea pentru ce se va întâmpla cu aceia care vor opta să nu se supună unei asemenea expansiuni a minții. Eu aș

scoate în evidență faptul că această inteligență suplimentară va rezida în esență în *nor* (rețeaua în expansiune exponențială de computere la care ne conectăm prin comunicarea online), unde este stocată acum cea mai mare parte a inteligenței mașinilor noastre. Când folosești un motor de căutare, când recunoști vorbirea cu telefonul, când consulți un asistent virtual cum ar fi Siri sau îți folosești telefonul pentru a traduce un termen într-o altă limbă, inteligența nu se află în aparatul în sine, ci în nor. Neocortexul nostru extins va fi și el găzduit tot acolo. Dacă accesăm o asemenea inteligență extinsă prin conexiuni neuronale directe, cum facem acum – interacționând cu ea prin intermediul dispozitivelor noastre –, operăm o distincție arbitrară. Din punctul meu de vedere, vom deveni mai creativi prin această amplificare atotprezentă, indiferent dacă alegem să ne conectăm sau nu la inteligența extinsă a omenirii. Deja am redirecționat o mare parte din memoria noastră personală, socială, istorică și culturală către nor și în cele din urmă vom face același lucru și cu gândirea noastră ierarhică.

Realizarea lui Einstein a rezultat nu numai din aplicarea metaforelor prin experimentele minții, ci și din curajul lui de a crede în forța acestor metafore. A fost capabil să părăsească explicațiile tradiționale care nu reușiseră să-i satisfacă experimentele și s-a arătat dispus să înfrunte ridiculizarea confraților lui la aflarea bizarelor explicații pe care le implicau metaforele lui. Aceste calități – credința în metafore și curajul convingerilor – sunt unele dintre cele pe care ar trebui să fim capabili să le programăm în neocortexul nostru nonbiologic.

Iubirea

> Limpezimea minții înseamnă și limpezimea pasiunii; acesta este motivul pentru care o minte mare și limpede iubește arzător și vede distinct ceea ce iubește.
>
> **— Blaise Pascal**

> Întotdeauna, iubirea are o anumită nebunie. Dar există întotdeauna și o anumită rațiune în nebunie.
>
> **— Friedrich Nietzsche**

Când vezi atât de multe în viață câte am văzut eu, nu mai subestimezi forța iubirii obsesive.

— Albus Dumbledore, în J.K. Rowling,
Harry Potter și prințul vitreg

Întotdeauna îmi place o soluție matematică bună la orice problemă de iubire.

— Michael Patrick King, în episodul intitulat
Du-mă la meci* din *Sex and the City

Dacă nu ai trăit cu adevărat și personal experiența iubirii extatice, fără îndoială că ai auzit de ea. Este corect să spun că o parte substanțială, dacă nu chiar majoritatea operelor de artă ale lumii – povești, romane, muzică, dans, picturi, emisiuni de televiziune și filme – sunt inspirate de povești de iubire în stadiile ei incipiente.

Recent, și știința a intrat în acest domeniu, iar acum suntem capabili să identificăm schimbările biochimice care se produc atunci când cineva se îndrăgostește. Se descarcă dopamină, care produce sentimente de fericire și desfătare. Nivelul sanguin al norepinefrinei crește foarte mult, ceea ce duce la intensificarea bătăilor inimii și la o stare generală de exaltare. Acești compuși chimici, împreună cu feniletilamina, produc jubilare, niveluri ridicate de energie, concentrarea atenției, pierderea poftei de mâncare și o sete arzătoare de a poseda obiectul dorințelor. Fapt interesant, cercetări recente de la University College din Londra arată, de asemenea, că nivelul serotoninei în sânge se reduce, situație similară cu cea care are loc în cazul tulburărilor obsesiv-compulsive, fapt ce corespunde naturii obsesive a iubirii incipiente.[2] Nivelurile ridicate ale dopaminei și norepinefrinei explică creșterea atenției pe termen scurt, euforia și dorința de a trăi iubirea incipientă.

Faptul că aceste fenomene biochimice par similare cu cele ale sindromului „Luptă sau fugi" este adevărat, doar că aici alergăm către ceva sau cineva; pe bună dreptate, cinicul ar putea să spună: *către* pericol, nu *departe* de el. De asemenea, schimbările corespund pe deplin cu cele ale fazelor incipiente ale comportamentului dependent. Piesa muzicală în stilul roxy music *Love Is the Drog* este cât se poate de elocventă în descrierea acestei stări (deși subiectul piesei îl constituie căutarea următoarei satisfaceri cu iubire). Studiile pe marginea experiențelor religioase extatice indică, de asemenea, aceleași fenomene fizice; se poate spune că persoana care

trăieşte o experienţă mistică se îndrăgosteşte de Dumnezeu sau de orice fel de legătură spirituală asupra căreia se concentrează.

În cazul iubirii romantice incipiente, estrogenii şi testosteronul joacă în mod sigur un rol în stabilirea pornirii sexuale, dar, dacă reproducerea sexuală ar fi singurul obiectiv evolutiv al iubirii, atunci aspectul romantic al procesului nu ar mai fi necesar. Aşa cum scria psihologul John William Money (1921-2006), „dorinţa e desfrânată, iubirea e lirică".

Faza extatică a iubirii conduce la faza de ataşament şi în cele din urmă la o legătură pe termen lung. Există compuşi chimici care încurajează şi acest proces, inclusiv ocitocina şi vasopresina. Să ne gândim la două specii înrudite de şoareci de câmp: şoarecele de câmp din prerie şi şoarecele de câmp de la munte. Sunt specii cât se poate de asemănătoare, cu excepţia faptului că şoarecele de prerie are receptori pentru ocitocină şi vasopresină, în timp ce şoarecele de munte nu îi are. Şoarecele de prerie este cunoscut pentru relaţiile sale monogame pe toată durata vieţii, în timp ce şoarecele de munte recurge aproape exclusiv la relaţii de o singură noapte. În cazul şoarecilor, ocitocina şi vasopresina au un rol determinant în privinţa naturii vieţii lor sexuale.

Deşi aceşti compuşi chimici exercită influenţe şi asupra omului, neocortexul nostru a preluat comanda, la fel ca în toate celelalte lucruri pe care le facem. Şoarecii de câmp au şi ei neocortex, dar cu dimensiuni de timbru şi plat. El este, totuşi, suficient de mare pentru ca şoarecii să îşi găsească o pereche pentru toată viaţa (sau, în cazul şoarecilor de munte, cel puţin de-o noapte) şi să-şi manifeste celelalte caracteristici elementare de şoareci. Noi, oamenii, avem suficient de mult neocortex adiţional ca să ne angajăm în expresiile expansive „lirice" la care se referă Money.

Dintr-o perspectivă evoluţionistă, chiar şi iubirea există pentru satisfacerea nevoilor neocortexului. Dacă nu am fi avut neocortex, atunci dorinţa carnală ar fi fost arhisuficientă pentru a ne garanta reproducerea. Provocarea extatică a iubirii conduce la ataşament şi dragoste matură şi se soldează cu o legătură durabilă. Aceasta, la rândul ei, este făcută să asigure cel puţin probabilitatea unui mediu stabil pentru copii în timpul în care neocortexul lor trece prin procesul critic de învăţare, necesar pentru a deveni adulţi responsabili şi capabili. Învăţarea într-un mediu bogat este o parte inerentă a metodei neocortexului. Aceleaşi mecanisme ale ocitocinei şi vasopresinei joacă un rol-cheie şi în stabilirea unei legături esenţiale între părinte (în special mamă) şi copil.

La capătul depărtat al poveștii iubirii, persoana iubită devine o parte importantă a neocortexului tău. După decenii petrecute împreună, în neocortex există un „celălalt" virtual, astfel încât putem să anticipăm pas cu pas ce va spune și va face el. Modelele noastre neocorticale sunt pline de gânduri și modele care reflectă cine este „celălalt". Când pierdem această persoană, literalmente pierdem o parte din noi înșine. Nu e doar o metaforă – toți recognizatorii modelului de ansamblu, care sunt plini de modele ce reflectă persoana pe care o iubim, își schimbă, brusc, natura. Cu toate că pot fi considerate un mijloc neprețuit de păstrare a celuilalt în viață înăuntrul nostru, vastele modele neocorticale ale unei persoane iubite pierdute devin brusc, din declanșatori ai încântării, declanșatori ai durerii.

Baza evolutivă a iubirii și fazele ei nu reprezintă totul în lumea de astăzi. În mare măsură, am reușit să eliberăm sexul de funcția sa biologică, prin faptul că putem face copii fără sex și, cu siguranță, putem face sex fără să facem copii. Vasta majoritate a raporturilor sexuale se produce în scopuri senzuale și relaționale. Și ne îndrăgostim, în mod curent, pentru alte motive decât cel de a face copii.

În mod similar, vasta cuprindere a expresiei artistice de toate genurile care celebrează iubirea și nenumăratele ei forme, datând încă din Antichitate, reprezintă de asemenea un scop de sine stătător. Și exact capacitatea noastră de a crea aceste forme durabile de cunoaștere transcendentală – referitoare la iubire sau la orice altceva – determină specia noastră să fie unică.

Neocortexul reprezintă cea mai mare creație a biologiei. Poeziile de iubire – și toate celelalte creații ale noastre – sunt cele mai mari creații ale neocortexului nostru.

Capitolul 7

NEOCORTEXUL DIGITAL INSPIRAT DE BIOLOGIE

Niciodată să nu te încrezi în ceva ce poate gândi independent dacă nu vezi unde îşi ţine creierul.

— Arthur Weasley, în J.K. Rowling,
Harry Potter şi prizonierul din Azkaban

Nu, nu mă interesează crearea unui creier puternic. Urmăresc doar un creier mediocru, cam ca al preşedintelui Companiei Americane de Telefonie şi Telegraf.

— Alan Turing

Un computer ar merita să fie considerat inteligent dacă ar putea înşela un om făcându-l să creadă că e om.

— Alan Turing

Cred că, la sfârşitul secolului, folosirea cuvintelor şi opinia generală avizată se vor modifica atât de mult, încât vei putea vorbi despre maşini care gândesc fără să te temi că ai putea fi contrazis.

— Alan Turing

Şobolanul-mamă construieşte un cuib pentru puii săi chiar dacă nu a mai văzut un alt şobolan în viaţa ei.[1] La fel, păianjenul îşi ţese pânza, coropişniţa îşi face propriul cocon, iar castorul construieşte un dig chiar dacă niciun contemporan nu le-a arătat vreodată cum se fac aceste lucruri. Asta nu înseamnă că aceste comportamente nu sunt învăţate, ci doar că aceste animale nu le-au învăţat în decursul unei singure vieţi, ci pe parcursul a mii de vieţi. Evoluţia comportamentului animalelor constituie pe bună dreptate un proces de învăţare, dar este vorba de învăţarea la nivelul speciei, nu al individului, iar rodul acestui proces de învăţare este codificat în ADN.

Pentru a înțelege semnificația evoluției neocortexului, să ne gândim că evoluția a accelerat foarte mult procesul de învățare (cunoaştere ierarhică) de la mii de ani la câteva luni (sau mai puțin). Chiar dacă milioane de animale dintr-o anumită specie de mamifere nu au reuşit să rezolve o problemă (necesitând o ierarhie de paşi), a fost nevoie ca doar unul singur să dea din întâmplare peste o soluție. Pe urmă, această nouă metodă a fost copiată şi răspândită exponențial în rândul populației.

Ne aflăm acum în poziția de a accelera din nou procesul de învățare de mii sau de milioane de ori, prin trecerea de la inteligența biologică la inteligența nonbiologică. Din momentul în care un neocortex digital învață o abilitate, poate să transfere acel know-how în câteva minute sau chiar secunde. Printre nenumăratele exemple: la prima mea companie, Kurzweil Computer Products (actualmente Nuance Speech Technologies), pe care am fondat-o în anul 1973, am avut nevoie de ani întregi pentru a pregăti un set de computere de cercetare ca să recunoască literele tipărite dintr-un document, tehnologie numită omnifont (orice tip de font), recunoaştere optică a caracterelor (*Optical Character Recognition* – *OCR*). Această tehnologie se află într-un continuu proces de dezvoltare de aproape patruzeci de ani, actualul produs numindu-se *OmniPage de la Nuance*. Dacă vrei ca ordinatorul tău să recunoască litere tipărite, nu trebuie să pierzi ani de-a rândul pentru a-l învăța acest lucru, aşa cum am făcut noi – poți pur şi simplu să descarci modelele evoluate deja învățate de computerele de cercetare sub formă de software. În anii 1980, am început programul de recunoaştere a vorbirii, iar acea tehnologie, care se află de asemenea într-un proces de continuă dezvoltare de câteva decenii, face parte din Siri. Şi din nou poți să descarci în câteva secunde modelele evoluate învățate de computerele de cercetare pe parcursul mai multor ani.

În cele din urmă, vom crea un neocortex artificial care să dispună de întreaga anvergură şi flexibilitate a corespondentului lui uman. Gândeşte-te numai la beneficii! Circuitele electronice sunt de milioane de ori mai rapide decât circuitele noastre biologice. La început, va trebui să dedicăm această sporire a vitezei, pe de-a-ntregul, compensării relativei lipse de paralelism din computerele noastre, dar în cele din urmă neocortexul digital va fi mult mai rapid decât varianta biologică şi va continua să-şi sporească viteza.

Când ne vom îmbunătăți propriul neocortex cu o versiune sintetică, nu va mai trebui să ne facem griji legate de cât de mult neocortex artificial suplimentar va avea loc fizic în corpul şi creierul nostru, pentru că, în cea mai

mare parte a lui, el se va afla în nor, ca majoritatea mijloacelor de computare pe care le folosim astăzi. Am estimat anterior că, la nivelul neocortexului nostru biologic, avem ceva de ordinul a 300 de milioane de recognizatori de modele. Doar atâta s-a putut înghesui în cutia noastră craniană, chiar şi cu inovaţia evoluţionistă a unei frunţi înalte şi în condiţiile în care neocortexul ocupă aproximativ 80 la sută din spaţiul disponibil. Din momentul în care începem să ne gândim la nor, nu va mai exista vreo limită naturală – vom putea să folosim miliarde sau mii de miliarde de recognizatori, în esenţă cât de mulţi ne-ar trebui şi cât de mulţi ne poate asigura, în fiecare moment, Legea accelerării veniturilor.

Ca să înveţe o nouă abilitate, un neocortex digital are nevoie în continuare de multe repetări ale lecţiilor, la fel ca neocortexul biologic, dar, din momentul în care un singur neocortex digital din vreun colţ al lumii învaţă ceva, el poate să le împărtăşească această cunoaştere tuturor celorlalte neocortexuri digitale, fără întârziere. Fiecare dintre noi îşi va putea avea propriile extensii de neocortex privat în nor, la fel cum astăzi beneficiem de propriile noastre stocuri de date personale.

În sfârşit, dar nu în cele din urmă, vom putea să salvăm o copie de rezervă a părţii digitale a inteligenţei noastre. După cum am văzut, nu e o simplă metaforă să afirmăm că în neocortexul nostru este stocată informaţie şi e înfiorător să ne gândim că nimic din această informaţie nu are salvată undeva o copie de rezervă. Desigur, există o cale prin care ne facem totuşi o asemenea copie de rezervă pentru o parte din informaţia din creier – scriind-o pe hârtie. Capacitatea de a transfera cel puţin o parte din gândirea noastră pe un mediu care poate supravieţui trupurilor noastre biologice a reprezentat un uriaş pas înainte, dar un volum foarte mare de date din creierul nostru continuă să rămână vulnerabil.

Simulările creierului

O abordare a construirii unui creier digital constă în simularea precisă a unuia biologic. De exemplu, David Dalrymple (născut în 1991), doctorand în ştiinţele creierului la Harvard, intenţionează să simuleze creierul unui nematod (vierme rotund).[2] Dalrymple a ales nematodul datorită sistemului său nervos relativ simplu, care însumează aproximativ 300 de neuroni şi pe care intenţionează să-l simuleze până la nivelul de detaliu al moleculelor. De

asemenea, cercetătorul va realiza o simulare computerizată a corpului său și a mediului, astfel încât nematodul lui virtual să poată vâna hrana (virtuală) și să poată face restul lucrurilor la care sunt buni nematozii. Dalrymple spune că e posibil să fie primul upload complet de creier între un animal biologic și unul virtual, care trăiește într-o lume virtuală. La fel ca în cazul nematodului simulat, încă se mai discută dacă nematozii biologici sunt conștienți, deși, în lupta lor pentru hrănire, digestie, evitarea prădătorilor și reproducere, au experiențe de care pot fi conștienți.

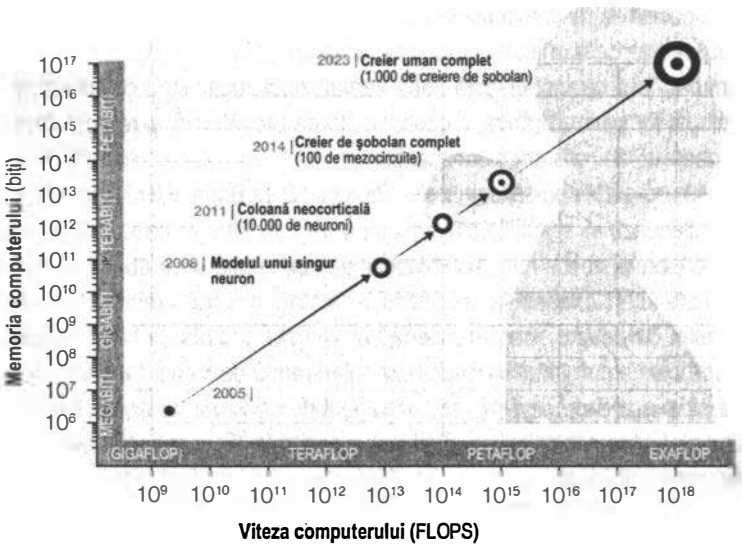
La capătul opus al spectrului, Proiectul Blue Brain al lui Henry Markram intenționează să simuleze creierul uman, inclusiv întregul neocortex și regiuni ale creierului vechi, cum ar fi hipocampul, amigdala și cerebelul. Simulările pe care intenționează să le construiască Markram vor fi realizate la diferite niveluri de detaliu, mergând până la o simulare totală, la nivel molecular. După cum am anunțat în Capitolul 4, Markram a descoperit un modul-cheie de câteva zeci de neuroni care se repetă de nenumărate ori la nivelul neocortexului, demonstrând că învățarea e realizată de aceste module și nu de neuronii individuali.

Progresele lui Markram se acumulează într-un ritm exponențial. A simulat un neuron în 2005, anul inițierii proiectului său. În anul 2008, echipa lui a simulat întreaga coloană neocorticală a unui creier de șobolan, care consta în 10.000 de neuroni. Până în anul 2011, a ajuns la un total de 100 de coloane, totalizând un milion de neuroni, pe care îi numește mezocircuit. O controversă privitoare la cercetările lui Markram ridică întrebarea cum se poate verifica dacă simulările sunt precise. Pentru a răspunde, simulările vor trebui să demonstreze că învață, așa cum voi arăta în cele ce urmează.

Markram intenționează să simuleze un creier de șobolan complet de 100 de mezocircuite, totalizând 100 de milioane de neuroni și în jur de trei miliarde de sinapse, până în anul 2014. În cursul unei discuții în cadrul conferinței TED din 2009 de la Oxford, Markram a spus: „Nu este imposibil de construit un creier uman și o putem fac peste zece ani".[3] Cea mai recentă țintă a lui în direcția simulării totale a creierului uman o reprezintă anul 2023.[4]

Markram și echipa lui își întemeiază modelul pe analize detaliate, anatomice și electrochimice, ale unor neuroni reali. Folosind un dispozitiv automat, pe care l-au denumit *robot patch-clamp*, ei măsoară canalele specifice de ioni, neurotransmițători și enzime care sunt răspunzătoare pentru activitatea electrochimică de la nivelul fiecărui neuron. Potrivit lui Markram, sistemul lor automat a reușit să realizeze în șase luni echivalentul a treizeci

de ani de analize. În cursul acestor analize, au fost observate unitățile de genul „memorie Lego", adică unitățile funcționale de bază ale neocortexului.

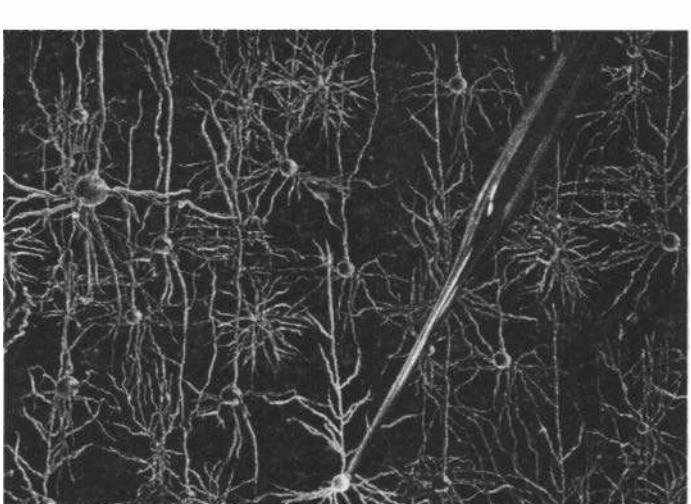

2023 | Creier uman complet
(1.000 de creiere de șobolan)
2014 | Creier de șobolan complet
(100 de mezocircuite)
2011 | Coloană neocorticală
(10.000 de neuroni)
2008 | Modelul unui singur neuron
2005 |
Memoria computerului (biți)
MEGABIȚI GIGABIȚI TERABIȚI PETABIȚI
GIGAFLOP TERAFLOP PETAFLOP EXAFLOP
Viteza computerului (FLOPS)

Progresele actuale şi anticipate ale Proiectului Blue Brain de simulare a creierului

Contribuții semnificative la tehnologia tehnicii *patch-clamping* robotizate au fost înregistrate de specialistul în neuroştiință de la MIT Ed Boyden, de profesorul de inginerie mecanică de la Georgia Tech Craig Forest şi de un absolvent al lui Forest, Suhasa Kodandaramaiah. Ei au efectuat demonstrații cu ajutorul unui sistem automat având precizia de un micrometru. Acesta poate să realizeze scanarea țesutului neuronal la distanțe foarte mici fără să lezeze membranele delicate ale neuronilor. „E un lucru pe care un robot îl poate face, dar omul, nu", a comentat Boyden.

Ca să revenim la Markram, după ce acesta a simulat coloana neocorticală, a fost citat ca spunând: „Acum mai trebuie doar să mărim la scară".[5] Mărirea la scară este în mod cert un factor important, dar mai există o problemă-cheie – şi anume învățarea. Dacă se doreşte ca creierul produs de Proiectul Blue Brain „să vorbească, să aibă inteligență şi să se comporte foarte asemănător unui om", aşa cum şi-a descris Markram obiectivul în cursul unui interviu acordat BBC în anul 2009, atunci va trebui ca acesta

să aibă în neocortexul său simulat un conținut suficient de mare pentru realizarea tuturor acestor funcții.[6] După cum poate atesta oricine a încercat să poarte o conversație cu un nou-născut, înainte ca acest lucru să devină fezabil, trebuie să învețe foarte mult.



Vârful robotului *patch-clamping* pus la punct la MIT și Georgia Tech, în timp ce scanează țesutul neuronal

Există două mijloace evidente prin care se poate atinge acest obiectiv în cazul unui creier simulat asemănător lui Blue Brain. Unul ar fi ca respectivul creier să învețe acest conținut așa cum o face un creier de om. La început, poate fi asemenea unui nou-născut cu o capacitate de acumulare a cunoașterii ierarhice înnăscută și cu anumite transformări preprogramate în regiunile senzoriale preprocesatoare. Însă învățarea ce are loc între un copil uman și un om care poate întreține o conversație ar trebui să se desfășoare într-o manieră comparabilă și în cazul învățării nonbiologice. Problema în cazul acestei abordări este că un creier simulat la nivelul de detaliu anticipat pentru Blue Brain e de așteptat să funcționeze în timp real cel mai curând la începutul anilor 2020. Chiar și funcționarea în timp real poate fi prea lentă dacă cercetătorii nu sunt pregătiți să aștepte un deceniu sau două pentru ca produsul lor să ajungă la paritate intelectuală cu un adult uman, deși

performanța în timp real va fi permanent mai bună, pe măsură ce computerele vor continua să evolueze sub aspectul raportului preț/performanță.

Cealaltă abordare presupune acțiunea de a se lua mai multe creiere umane biologice care au achiziționat deja suficientă cunoaștere pentru a conversa într-un limbaj inteligibil, pentru a se comporta în general de o manieră matură și pentru a-și copia modelele neocorticale în creierul simulat. În cazul acesta, problema e că metoda necesită o tehnologie de scanare noninvazivă și nondistructivă cu rezoluție spațială și temporală suficientă, dotată cu viteză pentru a îndeplini o asemenea sarcină rapid și complet. Nu m-aș aștepta ca o asemenea tehnologie de *upload* să fie disponibilă până în jurul anilor 2040. (Cerințele computaționale pentru simularea creierului cu o asemenea precizie, pe care o estimez la 1.019 calcule pe secundă, va fi accesibilă unui supercomputer, potrivit proiecțiilor mele, înspre începutul anilor 2020; însă tehnologiile nondistructive de scanare a creierului vor avea nevoie de mai mult timp.)

Există și o a treia abordare, la care, cred eu, ar putea să recurgă proiectele de simulare a creierului de tipul Blue Brain. Modelele moleculare se pot simplifica prin crearea unor echivalente funcționale la diferite niveluri de specificitate, mergând de la metoda mea funcțională algoritmică (pe care o descriu în această carte) până la simulări mai apropiate de simulările moleculare totale. Viteza de învățare poate fi, prin această tehnică, accelerată cu un ordin de mărime de sute sau mii, în funcție de gradul de simplificare folosit. Pentru creierul simulat, poate fi conceput un program educațional (cu ajutorul modelului funcțional) pe care acesta îl poate învăța relativ rapid. Pe urmă, întreaga simulare moleculară poate fi substituită cu modelul simplificat, care își folosește în continuare cunoștințele acumulate. Ulterior, putem simula învățarea cu modelul molecular complet, la o viteză mult mai redusă.

Specialistul american în știința computerelor Dharmendra Modha și colegii săi de la IBM au creat o simulare celulă cu celulă a unei porțiuni din neocortexul vizual uman cuprinzând 1,6 miliarde de neuroni virtuali și 9 mii de miliarde de sinapse, echivalentul unui neocortex de pisică. Această simulare funcționează de 100 de ori mai lent pe un supercomputer IBM BlueGene/P, format din 147.456 de procesoare. Cercetarea a primit Premiul Gordon Bell din partea Asociației pentru Mașini de Computare.

Scopul unui proiect de simulare a creierului precum Blue Brain și al unui neocortex simulat ca al lui Modha este în mod expres acela de a rafina și confirma un model funcțional. Inteligența artificială la nivel uman va folosi,

în principal, modelul funcțional algoritmic discutat în această carte. Pe de altă parte, simulările moleculare ne vor ajuta să perfecționăm acest model și să înțelegem pe deplin ce detalii sunt importante. În cursul cercetărilor noastre privitoare la punerea la punct a tehnologiei de recunoaștere a vorbirii din anii 1980 și 1990, am reușit să ne rafinăm algoritmii din momentul în care am înțeles transformările propriu-zise realizate de nervul auditiv și de porțiunile timpurii ale cortexului auditiv. Deși modelul nostru funcțional a fost perfect, abia atunci când vom înțelege cu precizie cum este el implementat practic în creierul nostru biologic, ni se vor dezvălui cunoștințele importante referitoare la funcționalitatea și disfuncționalitatea umană.

Vom avea nevoie de date detaliate referitoare la creierul real pentru a realiza simulări bazate pe biologie. Echipa lui Markram își adună propriile date. Există proiecte pe scară largă de colectare a acestor date și de punere a lor la dispoziția oamenilor de știință. De exemplu, Laboratorul Cold Spring Harbor din New York a strâns, în urma scanării unui creier de mamifer (șoarece), 500 de terabiți de informații, pe care le-a făcut accesibile în luna iunie 2012. Proiectul îi permite utilizatorului să exploreze un creier tot așa cum Google Earth permite explorarea suprafeței planetei. Te poți deplasa în jurul întregului creier mărindu-l pentru a vedea neuronii individuali și conexiunile lor. Poți scoate în evidență o singură conexiune și pe urmă îi poți urmări traseul prin creier.

Șaisprezece secții ale Institutelor Naționale pentru Sănătate s-au reunit și au sponsorizat o inițiativă majoră, intitulată Human Genome Project – Proiectul Genomului Uman –, cu suma de 38,5 milioane de dolari.[7] Derulat de Universitatea Washington din St. Louis, Universitatea din Minnesota, Universitatea Harvard, Spitalul General Massachusetts și Universitatea California din Los Angeles, proiectul își propune să creeze o hartă similară tridimensională a conexiunilor din creierul uman. El folosește o varietate de tehnologii de scanare noninvazive, inclusiv noi forme de RMN, magnetoencefalografie (măsurarea câmpurilor magnetice produse de activitatea electrică a creierului) și tractografie prin difuziune (metodă de trasare a căilor fasciculelor de fibre din creier). După cum subliniez în Capitolul 10, rezoluția spațială a scanării noninvazive a creierului se îmbunătățește cu o rată exponențială. Cercetările lui Van J. Wedeen și ale colegilor săi de la Spitalul General Massachusetts, care au relevat o structură în grilă extrem de regulată a cablării neocortexului (vezi Capitolul 4), reprezintă un rezultat timpuriu al acestui proiect.

Specialistul în neuroştiinţă computaţională Anders Sandberg (născut în anul 1972) de la Universitatea Oxford şi filosoful suedez Nick Bostrom (născut în anul 1973) au scris o lucrare cuprinzătoare intitulată *Whole Brain Emulation: A Roadmap – Imitarea creierului în ansamblu: îndrumar*, care detaliază

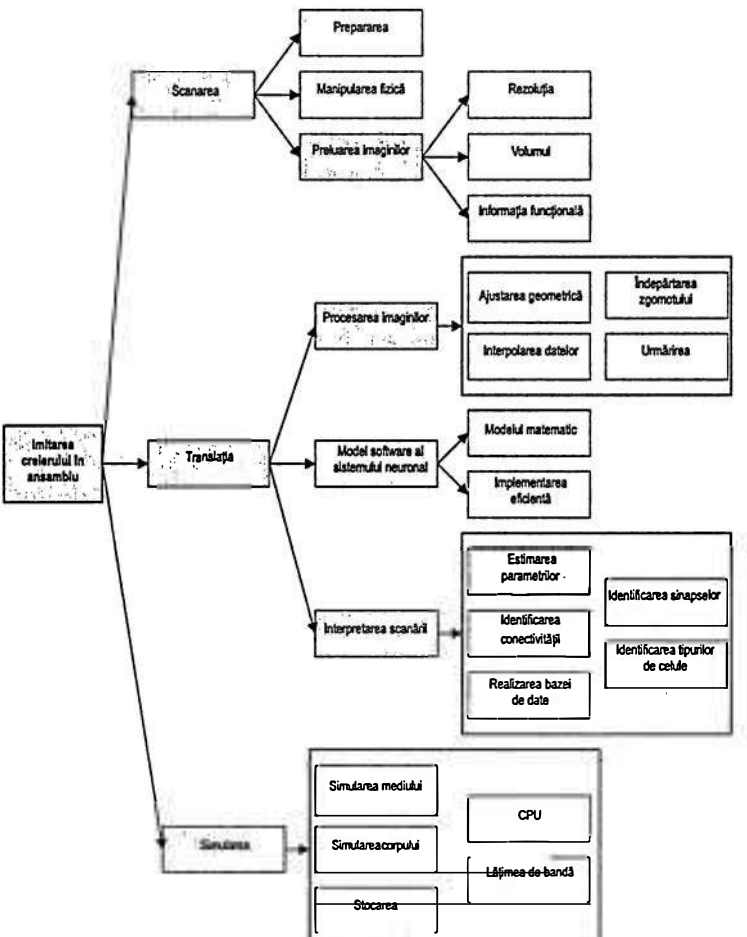
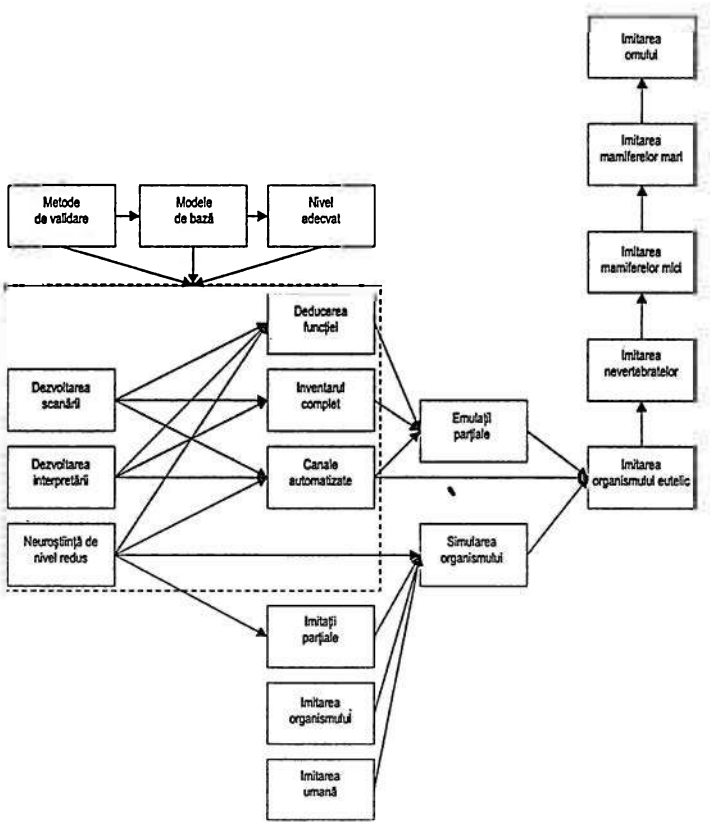


Diagrama capacităţilor tehnologice necesare pentru imitarea creierului în ansamblul său, din *Whole Brain Emulation: A Roadmap – Imitarea creierului în ansamblu: îndrumar*, de Anders Sandberg şi Nick Bostrom

cerințele referitoare la imitarea creierului uman (şi a altor tipuri de creiere) pe diferite niveluri de specificitate, de la modelele funcționale de nivel superior, la simularea moleculelor.[8] Comunicarea nu oferă o cronologie, însă descrie cerințele necesare pentru simularea diferitelor tipuri de creier la diferite niveluri de precizie sub aspectul scanării cerebrale, realizării modelelor, stocării şi computării. Comunicarea proiectează evoluții exponențiale continue în toate aceste domenii ale capabilității şi susține că exigențele privind simularea creierului uman la un nivel superior de detaliu sunt adecvate.



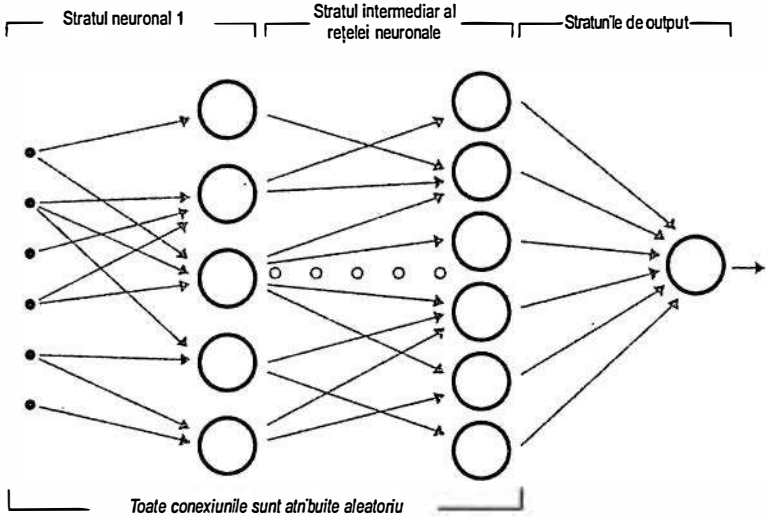
Rezumatul cărții *Whole Brain Emulation: A Roadmap – Imitarea creierului în ansamblu: îndrumar*, de Anders Sandberg şi Nick Bostrom

Rețele neuronale

În anul 1964, la vârsta de şaisprezece ani, i-am scris lui Frank Rosenblatt (1928-1971), profesor la Universitatea Cornell, adresându-i întrebări cu privire la o maşină denumită Mark 1 Perceptron. O crease cu patru ani înainte şi era descrisă ca având proprietăți asemănătoare creierului. El m-a invitat să-l vizitez şi să încerc maşina.

Perceptronul era construit din ceea ce Rosenblatt pretindea că erau modele electronice ale neuronilor. Input-ul consta în valori aranjate în două dimensiuni. Pentru vorbire, o dimensiune reprezenta frecvența iar cealaltă timpul, astfel încât fiecare valoare ilustra intensitatea unei frecvențe într-un anumit moment. Pentru imagini, fiecare punct era un pixel dintr-o imagine bidimensională. Fiece punct de un anumit input era conectat aleatoriu la input-urile primului strat de neuroni simulați. Fiecare conexiune avea asociată o tărie sinaptică pe măsura importanței ei, iar aceasta era stabilită inițial la o valoare aleatorie. Fiece neuron aduna semnalele care ajungeau la el. În caz că semnalul combinat depăşea un anumit prag, neuronul emitea un semnal la conexiunea sa de output; dacă semnalul combinat nu depăşea pragul, neuronul nu emitea semnalul, aşa că output-ul lui era zero. Output-ul fiecărui neuron era conectat aleatoriu la input-urile neuronilor din următorul strat. Mark 1 Perceptron avea trei straturi, care puteau fi organizate într-o largă varietate de configurații. De exemplu, un strat putea să transmită feedback la alt strat anterior. În stratul superior, output-ul unuia sau al mai multor neuroni, de asemenea aleşi aleatoriu, asigura răspunsul. (Pentru o descriere algoritmică a rețelelor neuronale, vezi această notă de final.)[9]

Deoarece deconectarea rețelei neuronale şi greutățile sinaptice sunt stabilite la început aleatoriu, răspunsurile unei asemenea rețele fără pregătire vor fi şi ele aleatorii. Prin urmare, rețeaua neuronală trebuie să-şi învețe materia, exact cum o face creierul mamiferelor, al cărui model se presupune că l-ar reprezenta. Rețeaua neuronală debutează într-o stare de ignoranță; profesorul – care poate fi un om, un program de computer sau o rețea neuronală mai matură, care şi-a învățat deja lecțiile – recompensează rețeaua-student atunci când aceasta generează output-ul corect şi o pedepseşte când nu o face. Acest feedback este folosit la rândul lui de către rețeaua neuronală student pentru ajustarea tăriei fiecărei conexiuni


Stratul neuronal 1
Stratul intermediar al rețelei neuronale
Stratul de output
Toate conexiunile sunt atribuite aleatoriu

interneuronale. Conexiunile care concordă cu răspunsul corect sunt întărite. Cele care dau răspunsuri greșite sunt slăbite.

Cu timpul, rețeaua neuronală se organizează pentru a da răspunsurile corecte fără pregătire. Experimentele au arătat că rețelele neuronale își pot învăța materia chiar și cu profesori care nu prezintă garanții. Chiar dacă dascălul are dreptate în numai 60 la sută dintre cazuri, rețeaua neuronală student își învață lecțiile cu o acuratețe ce se apropie de 100 la sută.

Pe de altă parte însă, au devenit evidente limitările în privința materialului pe care Perceptronul era capabil să-l învețe repede. Când l-am vizitat pe profesorul Rosenblatt, în anul 1964, am încercat modificări simple ale input-ului. Sistemul era setat să recunoască litere tipărite și putea să le recunoască precis. Se descurca destul de bine la autoasociere (respectiv, putea să recunoască literele chiar dacă erau acoperite părți din ele), dar se descurca mai puțin bine cu invarianța (respectiv generalizarea indiferent de mărime și font, care îi provocau confuzii).

În cursul ultimei jumătăți a anilor 1960, aceste rețele neuronale au devenit enorm de populare, iar domeniul „conexionismului" a pus stăpânire pe mai mult de jumătate din domeniul inteligenței artificiale. Între timp, o abordare mai tradițională a inteligenței artificiale a inclus tentative de

programare a soluțiilor pentru probleme specifice, cum ar fi recunoaşterea proprietăților invariabile ale literelor tipărite.

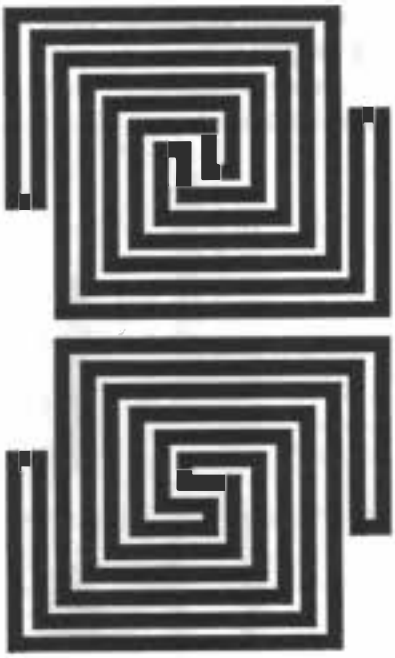
O altă persoană pe care am vizitat-o în anul 1964 a fost Marvin Minsky (născut în anul 1927), unul dintre fondatorii domeniului inteligenței artificiale. În ciuda faptului că, în anii 1950, făcuse personal o serie de cercetări de pionierat în domeniul rețelelor neuronale, era îngrijorat de marele val de interes față de această tehnică. Parte din farmecul rețelelor neuronale consta în faptul că se presupunea că acestea nu necesitau programare – că puteau să învețe soluțiile la probleme pe cont propriu. În anul 1965 am intrat la MIT, avându-l pe profesorul Minsky drept mentor, şi i-am împărtăşit scepticismul în legătură cu nebunia „conexionismului".

În anul 1969, Minsky şi Seymour Papert (născut în anul 1928), cofondatori ai Laboratorului de Inteligență Artificială de la MIT, au scris o carte intitulată *Percetrons – Perceptroni*, care prezenta o singură teoremă centrală şi anume că Perceptronul era în mod inerent incapabil să determine faptul că o imagine era conectată sau nu. Cartea a provocat un uragan. Determinarea faptului că o imagine este conectată sau ba reprezintă o sarcină pe care oamenii o pot îndeplini cu uşurință; este, de asemenea, un proces simplu să programezi un computer pentru a face această diferențiere. Faptul că Perceptronul nu putea să realizeze acest lucru a fost considerat de mulți un neajuns fatal.

Oricum, *Perceptrons – Perceptroni* a fost interpretată pe scară largă ca implicând mai multe decât o făcea în realitate. Teorema lui Minsky şi Papert se aplica doar unui tip particular de rețea neuronală, denumit feedforward (categorie care include şi Perceptronul lui Rosenblatt); alte tipuri de rețele neuronale nu aveau această limitare. Totuşi, cartea a reuşit să ucidă în mare măsură finanțarea cercetărilor pe tema rețelelor neuronale în cursul anilor 1970. Domeniul a revenit însă în anii 1980 cu tentative de folosire a unor modele mai realiste ale neuronilor biologici şi a unora care evitau limitările pe care le implica Teorema Perceptron a lui Minsky-Papert. Cu toate acestea, abilitatea neocortexului de a rezolva problema invarianței, cheia forței sale, a rămas în continuare insolvabilă pentru conexioniştii renăscuți ai domeniului.

Codificarea sporadică: cuantificarea vectorilor

La începutul anilor 1980, am demarat un proiect dedicat unei alte probleme clasice de recunoaştere a modelelor: înțelegerea vorbirii umane. La

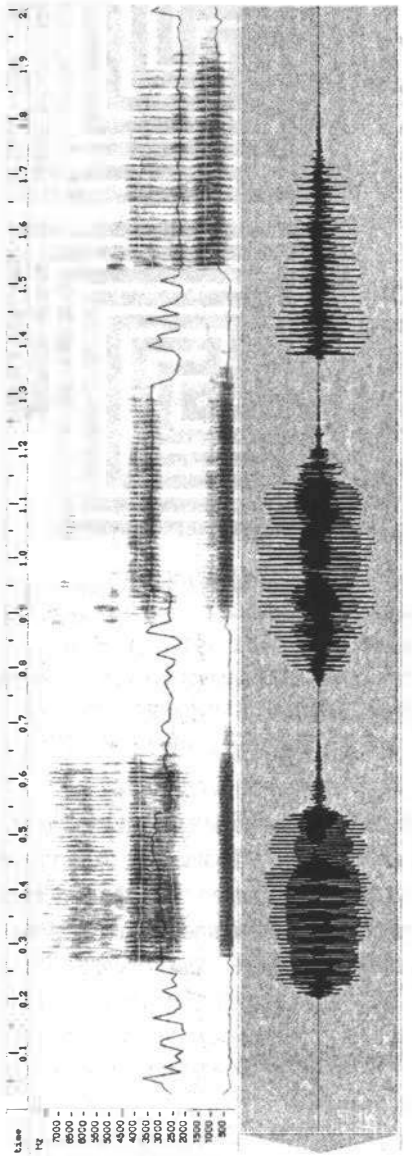


Două imagini de pe coperta cărții *Perceptrons – Perceptroni* a lui Marvin Minsky şi Seymour Papert. Imaginea de sus nu este conectată (respectiv, zonele întunecate constau din două părți deconectate). Imaginea de jos e conectată. Omul poate determina cu uşurință acest lucru, la fel ca un program software simplu. Un Perceptron cu feedforward, cum era Perceptronul Mark 1 al lui Frank Rosenblatt, nu putea să facă această determinare

început, am folosit abordări tradiționale pentru inteligența artificială prin programarea directă a cunoştințelor la nivel de expert referitoare la unitățile fundamentale de vorbire – fonemele – şi la regulile lingvistice privitoare la modul în care oamenii înşiruiesc fonemele pentru a forma cuvinte şi expresii. Fiecare fonem are modele distincte de frecvențe. De exemplu, ştiam că vocale precum „e" şi „a" sunt caracterizate de anumite frecvențe rezonante denumite formatori, cu o proporție caracteristică de formatori pentru fiecare fonem. Sunetele sibilante „z" şi „s" sunt definite de o explozie sonoră care cuprinde mai multe frecvențe.

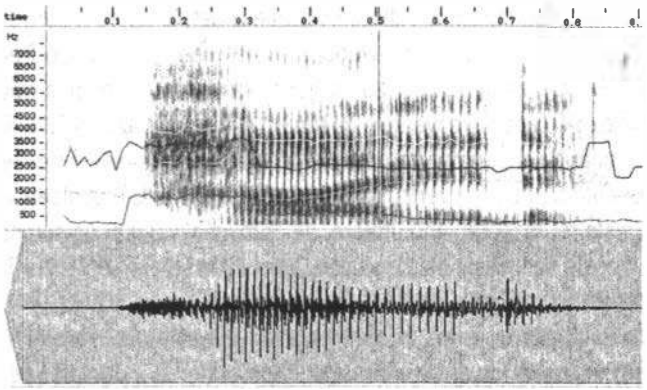
Am captat vorbirea sub formă de undă, pe care am convertit-o în mai multe benzi de frecvență (percepute ca înălțime), folosind o bancă de filtre

de frecvenţe. Rezultatul acestei transformări putea fi vizualizat şi a fost denumit spectrogramă.



Spectrograma pentru trei vocale. De la stânga la dreapta: [i] ca în *„appreciate"*, [u] ca în *„acoustic"* şi [a] ca în *„ah"*. Axa *y* reprezintă frecvenţa sunetului. Cu cât banda e mai întunecată, cu atât energia acustică pe acea frecvenţă este mai mare

Banca de filtre copiază ceea ce face cohleea umană, acesta fiind pasul inițial în procesarea biologică a sunetului. Programul software a identificat la început fonemele pe baza distincției dintre modelele de frecvență și pe urmă a recunoscut cuvinte pe baza identificării succesiunilor caracteristice de foneme.



Spectrogramă a unei persoane care rostește cuvântul „hide". Liniile orizontale arată formatorii, care sunt frecvențe susținute, cu o energie deosebit de mare[10]

Rezultatul a fost parțial încununat de succes. Puteam să ne învățăm dispozitivul să învețe modele pentru o anumită persoană care folosea un vocabular de dimensiuni moderate, măsurat în câteva mii de cuvinte. Când am încercat să recunoaștem zeci de mii de cuvinte, să lucrăm cu mai mulți vorbitori și să permitem vorbirea continuă (respectiv vorbirea fără pauze între cuvinte), am dat peste problema invarianței. Diferiți oameni pronunță aceleași foneme diferit – de exemplu, fonemul „e" al unei persoane poate să semene cu fonemul „ah" al alteia. Chiar și una și aceeași persoană prezintă inconsecvență în modul de rostire a unui anumit fonem. Modelul unui fonem a fost adeseori afectat de alte foneme învecinate. Multe foneme au fost eliminate complet. Pronunția cuvintelor (respectiv felul în care sunt înșirate fonemele pentru formarea cuvintelor) era de asemenea extrem de variabilă și dependentă de context. Regulile lingvistice pe care le programasem se prăbușeau și nu puteau ține pasul cu extrema variabilitate a limbajului vorbit.

La vremea respectivă, mi-a devenit clar că esența modelului uman și a recogniției conceptuale se bazează pe ierarhii. Acest fapt este cu certitudine evident pentru limbajul uman, care constituie o ierarhie elaborată de structuri. Dar care este elementul de la baza structurilor? Aceasta a fost prima întrebare pe care am avut-o în vedere atunci când am căutat mijloacele de recunoaștere automată a vorbirii umane integrale.

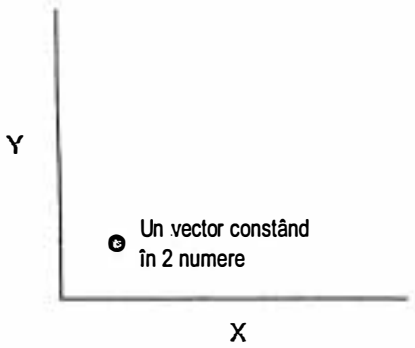
Sunetul pătrunde în ureche ca vibrație a aerului și este convertit, de aproximativ 3.000 de celule cu perișori din cohlee, în benzi de frecvență multiple. Fiecare celulă cu perișori e acordată pentru o anumită frecvență (observă că percepem frecvențele ca tonuri) și funcționează ca filtru de frecvență, emițând un semnal ori de câte ori există un sunet cu frecvența ei de rezonanță sau cu frecvență apropiată. Când părăsește cohleea umană, sunetul este reprezentat, prin urmare, de aproximativ 3.000 de semnale separate, fiecare semnificând intensitatea variabilă în timp a unei benzi înguste de frecvență (cu o suprapunere substanțială între benzi).

Cu toate că devenise evident că creierul era masiv paralel, mie mi s-a părut imposibil ca el să efectueze împerecherea modelelor pe 3.000 de semnale auditive separate. M-am îndoit că evoluția a putut fi într-atât de eficientă. Actualmente știm că, într-adevăr, se realizează o reducere substanțială a datelor la nivelul nervului auditiv înainte ca semnalele să ajungă la cortex.

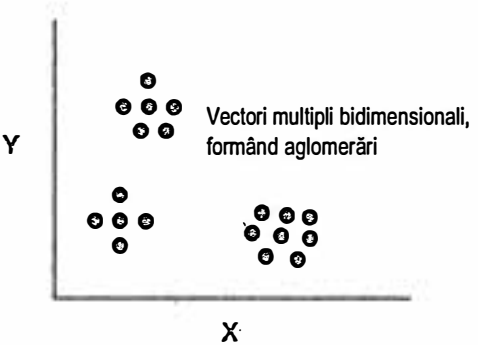
La recognizatorii noștri de vorbire bazați pe sofware, am folosit de asemenea filtre implementate ca software – șaisprezece, ca să fiu mai exact (număr pe care ulterior l-am mărit la treizeci și doi, deoarece am constatat că nu exista niciun beneficiu dacă depășeam acest număr). Prin urmare, în sistemul nostru, fiecare punct din timp era reprezentat de șaisprezece numere. Am fost obligați să reducem aceste șaisprezece fluxuri de date la unul singur și în același timp să accentuăm caracteristicile semnificative în recunoașterea vorbirii.

Noi am folosit o tehnică matematică optimă, numită cuantificarea vectorială. Să spunem că, în orice moment de timp, sunetul (cel puțin de la o ureche) a fost reprezentat de software-ul nostru prin șaisprezece numere diferite: respectiv output-ul celor șaisprezece filtre de frecvență. (La nivelul sistemului auditiv uman, totalul ar fi de aproximativ 3.000, cifră reprezentând output-ul celor 3.000 de celule interioare cohleare cu perișori.) În terminologie matematică, fiecare asemenea set de numere (indiferent că ar fi 3.000 în cazul biologic sau 16 în implementarea noastră software) se numește *vector*.

Pentru simplificare, să luăm procesul cuantificării cu vectori pentru două numere. Fiecare vector poate fi considerat un punct la nivel bidimensional.

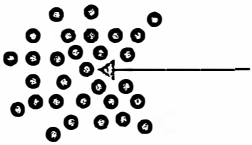


Dacă avem o mostră foarte mare de asemenea vectori şi îi plotăm, este posibil să observăm formarea de aglomerări



În scopul identificării aglomerărilor, trebuie să decidem cât de multe vom accepta. În cadrul proiectului nostru, am acceptat în general 1.024 de aglomerări, astfel încât să le putem număra şi să putem atribui fiecărei aglomerări o etichetă de 10 biți (pentru că 210 = 1.024). Mostra noastră de vectori ilustrează diversitatea pe care o aşteptam. Am hotărât cu titlu de încercare ca primii 1.024 de vectori să fie aglomerări de puncte individuale. Pe urmă am luat vectorul cu numărul 1.025 şi am găsit punctul de care era cel mai aproape. Dacă această distanță este mai mare decât distanța cea mai

mică dintre orice pereche de 1.024 puncte, considerăm că ea este punctul de început al unei noi aglomerări. Pe urmă, reducem cele mai apropiate două aglomerări (de puncte individuale) la o singură aglomerare. În acest fel, tot ne-au mai rămas 1.024 de aglomerări. După procesarea vectorului cu numărul 1.025, una dintre aceste aglomerări are de-acum mai mult de un singur punct. Continuăm să procesăm punctele în acest fel, păstrând în permanență 1.024 de aglomerări. După ce am procesat toate punctele, reprezentăm fiecare aglomerare multipunct prin centrul geometric al punctelor din aglomerarea respectivă.



O aglomerare de puncte. O reprezentăm folosind un singur punct, care este centrul geometric al tuturor punctelor aglomerării

Continuăm acest proces iterativ până când folosim toate punctele mostrei. De obicei, procesăm un milion de puncte în aglomerări de câte 1.024 (210); am mai folosit şi aglomerări de 2.048 (211) sau 4.096 (212). Fiecare aglomerare este reprezentată de un vector situat în centrul geometric al tuturor punctelor din aglomerarea respectivă. Astfel, totalul distanțelor tuturor punctelor din aglomerare față de punctul central al aglomerării este cel mai mic posibil.

Rezultatul acestei tehnici este că, în loc să existe un milion de puncte de la care am început (şi un număr mai mare chiar de posibile puncte), am redus datele la numai 1.024 de puncte, care folosesc în mod optim spațiul posibilităților. Părților aparținând acestui spațiu care nu sunt folosite niciodată nu li se atribuie vreo aglomerare.

Pe urmă, atribuim un număr fiecărei aglomerări (în cazul nostru, de la 0 la 1.023). Acest număr este reprezentarea redusă, „cuantificată" a acelei aglomerări, motiv pentru care tehnica se numeşte cuantificare vectorială. Orice nou input care va sosi în viitor va fi reprezentat prin numărul aglomerării al cărei centru este cel mai apropiat de acest nou vector de input.

Acum, putem precomputa un tabel cu distanțele de la punctele centrale ale fiecărei aglomerări până la fiecare dintre celelalte puncte centrale. În acest mod, avem instantaneu la dispoziție distanța noului vector de input

(pe care îl reprezentăm prin acest punct cuantificat – cu alte cuvinte, prin numărul aglomerării de al cărei centru este cel mai apropiat punctul) pentru fiecare aglomerare. Deoarece reprezentăm punctele doar prin aglomerările cele mai apropiate de ele, cunoaştem de-acum distanța de la acest punct până la orice alt punct care ar putea să apară.

Am descris tehnica de mai înainte folosind vectori cu numai două numere fiecare, dar este absolut analog cu exemplul mai simplu dacă se lucrează cu vectori cu şaisprezece elemente. Pentru că am ales vectori cu şaisprezece numere reprezentând şaisprezece benzi diferite, fiecare punct din sistemul nostru are un punct în spațiul cu şaisprezece dimensiuni. Nouă ne este greu să ne imaginăm un spațiu cu mai mult de trei dimensiuni (poate patru, dacă includem şi timpul), dar matematica nu are asemenea inhibiții.

Prin intermediul acestui proces, am realizat patru lucruri. Primul, am redus foarte mult complexitatea datelor. Al doilea, am redus datele în şaisprezece dimensiuni la date unidimensionale (respectiv, fiecare mostră este acum un singur număr). Al treilea, ne-am îmbunătățit capacitatea de a găsi caracteristici invariabile, deoarece scoatem în evidență porțiuni ale spațiului posibilelor sunete care transmit cele mai multe informații. Cea mai mare parte a combinațiilor de frecvențe e fizic imposibilă sau foarte puțin probabilă, astfel încât nu există niciun motiv pentru care să se aloce unor combinații improbabile de input-uri un spațiu egal cu acela pentru combinațiile probabile. Această tehnică reduce datele la posibilități cu aceeaşi probabilitate. Al patrulea beneficiu este acela că putem folosi recognizatori de modele unidimensionali, chiar dacă datele originale constau în mult mai multe dimensiuni. Aceasta s-a dovedit a fi abordarea cea mai eficientă pentru utilizarea resurselor disponibile de computație.

Citirea gândurilor cu Modele Markov Ascunse

Prin cuantificarea vectorială, am simplificat datele într-un mod ce pune accentul pe caracteristicile-cheie, însă tot ne mai trebuia un mod de reprezentare a ierarhiei caracteristicilor invariabile care să facă logică noua informație. Pentru că în acel moment lucrasem deja în domeniul recognizatorilor de modele timp de douăzeci de ani, ştiam că reprezentările unidimensionale erau mult mai puternice, eficiente şi pretabile la rezultate invariabile. Neocortexul nu era prea bine cunoscut la începutul anilor 1980, dar, pe baza

experienței pe care o aveam cu o multitudine de probleme de recunoaştere, am presupus că era posibil ca şi creierul să reducă datele multidimensionale (provenite fie de la ochi, fie de la urechi sau piele) folosind o reprezentare unidimensională, mai ales că se dezvoltau conceptele privitoare la ierarhizarea neocortexului.

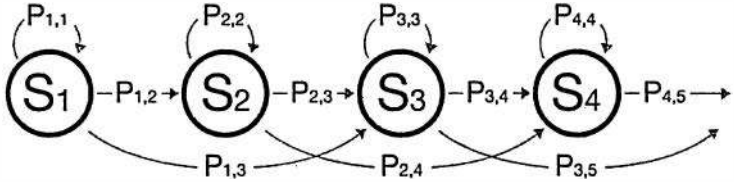
Pentru problema recunoaşterii vorbirii, organizarea informației din semnalul vorbit a părut să însemne o ierarhie de modele, fiecare model fiind reprezentat de un şir liniar de elemente, cu o direcție de înaintare. Fiece element al modelului putea fi un alt model de la un nivel inferior sau o unitate fundamentală de input (care, în cazul recunoaşterii vorbirii, ar fi vectorii noştri cuantificați).

Îți vei da seama, cititorule, că această situație este în concordanță cu modelul neocortexului pe care ți l-am prezentat mai înainte. Vorbirea umană, aşadar, este produsă de o ierarhie de modele liniare din creier. Dacă am putea să examinăm pur şi simplu aceste modele din creierul persoanei care vorbeşte, totul s-ar rezuma la o simplă problemă de împerechere a noilor emisii verbale cu modelele cerebrale şi s-ar putea înțelege ce vrea să spună persoana respectivă. Din păcate, nu avem acces direct la creierul vorbitorului – singura informație pe care o obținem fiind ceea ce spune el. Desigur, exact acesta este şi scopul limbajului vorbit – vorbitorul îşi împărtăşeşte nişte idei prin vorbire.

Şi atunci m-am întrebat: există vreo tehnică matematică prin care e posibil să deducem modelele din creierul vorbitorului pe baza cuvintelor rostite? Desigur, o singură repriză vorbită nu ar fi suficientă, dar, dacă am avea un număr mare de mostre, am putea să folosim această informație pentru a citi practic modelele din neocortexul vorbitorului (sau cel puțin să formulăm ceva echivalent matematic care să ne permită să recunoaştem noile reprize de vorbire)?

Oamenii nu reuşesc să îşi dea seama cât de minunată poate fi matematica – reține, cititorule, că posibilitatea noastră de a căuta mare parte din cunoaşterea umană într-o fracțiune de secundă folosind motoarele de căutare se bazează pe o tehnică matematică. În cazul problemei legate de recunoaşterea vorbirii, cu care mă confruntam la începutul anilor 1980, s-a dovedit că Modelele Ascunse ale lui Markov răspundeau cât se poate de bine cerințelor. Matematicianul rus Andrei Andreeevici Markov (1856-1922) a construit o teorie matematică a succesiunilor ierarhice de stări. Modelul se baza pe posibilitatea parcurgerii stărilor într-un singur şir, iar, dacă acest

lucru se realiza cu succes, pe generarea unei stări pe următorul nivel superior al ierarhiei. Sună familiar?



Un exemplu simplu de strat al unui Model Markov Ascuns. Cu S_1 până la S_4 am notat stările interne „ascunse". Tranzițiile $P_{i,j}$ reprezintă, fiecare, probabilitatea de trecere de la starea S_i la starea S_j. Aceste probabilități sunt determinate de sistemul care învață din datele de pregătire (inclusiv în cursul utilizării propriu-zise). O nouă succesiune (ca de pildă o nouă repriză de vorbire) este comparată cu aceste probabilități pentru a se determina probabilitatea ca modelul să producă succesiunea

Modelul lui Markov include probabilități de apariție cu succes a fiecăreia dintre aceste stări. El a mers şi mai departe şi a formulat ipoteza unei situații în care un sistem are o asemenea ierarhie de succesiuni liniare de stări, dar ele nu pot fi examinate direct – de unde şi denumirea de Modele Markov Ascunse. Nivelul cel mai de jos al ierarhiei emite semnale, acestea reprezentând tot ce ne este permis să vedem. Markov a furnizat o tehnică matematică de computare a probabilităților pentru fiecare tranziție pe baza output-ului observat. Metoda a fost ulterior rafinată de Norbert Wiener, în anul 1923. Perfecționarea lui Wiener a asigurat, de asemenea, un mijloc de determinare a conexiunilor pentru Modelul Markov; în esență, orice conexiune cu o probabilitate prea redusă era considerată inexistentă. În esență, acesta este modul în care taie neocortexul uman conexiunile – dacă sunt rar sau niciodată folosite, sunt considerate improbabile şi sunt retezate. În cazul nostru, output-ul observat este semnalul verbal creat de persoana care vorbeşte, iar probabilitățile de stare şi conexiunile Modelului Markov constituie ierarhia neocorticală care l-a produs.

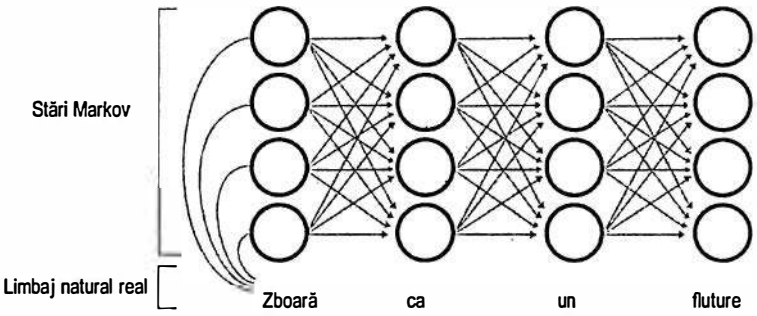
Eu am imaginat un sistem prin intermediul căruia să luăm mostre verbale umane, să le aplicăm tehnica Modelului Markov Ascuns pentru determinarea unei ierarhii de stări cu conexiuni şi probabilități (în esență, un neocortex simulat pentru producerea vorbirii) şi apoi să folosim această rețea

ierarhică determinată de stări pentru recunoaşterea noilor input-uri verbale. Pentru crearea unui sistem independent de vorbitor, puteam folosi mostre de la mai multe persoane diferite, pentru antrenarea Modelelor Markov. Adăugând şi elementul ierarhizării pentru reprezentarea naturii ierarhice a informaţiei în limbă, le-am numit în consecinţă Modele Markov Ascunse Ierarhice, MMAI (*Hierarchical Hidden Markov Models* – HHMM).

Colegii mei din compania Kurzweil Applied Intelligence au fost sceptici în legătură cu funcţionarea acestei metode, dat fiind faptul că era o metodă autoorganizatoare, reminiscentă a reţelelor neuronale, care nu mai erau la modă şi cu care avusesem prea puţin succes. Eu am scos în evidenţă faptul că reţeaua unui sistem neuronal în reţea este fixă şi nu se adaptează input-ului: greutăţile se adaptează, dar nu şi conexiunile. În cazul sistemului bazat pe Modelul Markov, dacă acesta era conceput corect, sistemul urma să taie conexiunile nefolosite astfel încât, în esenţă, să adapteze topologia.

Am creat ceea ce se considera a fi un proiect „skunk works" (termen organizaţional aplicat unui proiect care se abate de la calea clasică şi dispune de puţine resurse formale), compus din mine, un programator part-time şi un inginer din domeniul electric (pentru crearea băncii de filtre de frecvenţe). Spre surpriza colegilor mei, efortul nostru a fost încununat de succes, sistemul reuşind să recunoască vorbirea pentru un vocabular larg, cu o mare acurateţe.

După acel experiment, toate eforturile noastre în domeniul recunoaşterii vorbirii s-au bazat pe Modelele Markov Ascunse Ierarhice. Se pare că şi alte



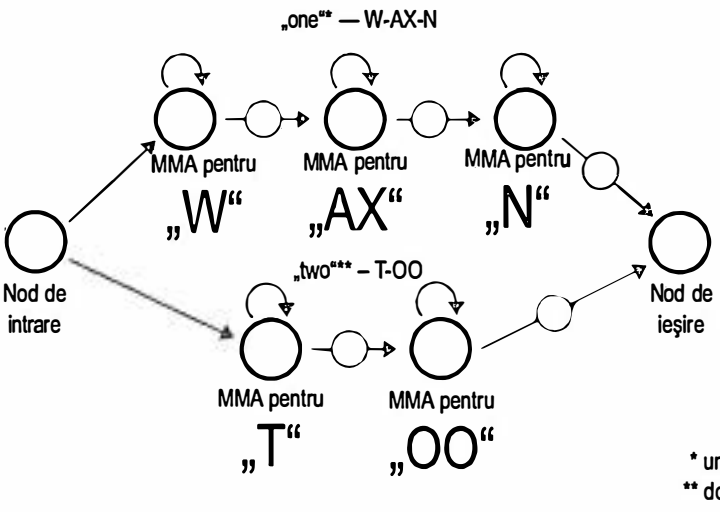
Stări Markov Ascunse şi posibile tranziţii pentru producerea unei succesiuni de cuvinte în text în limbaj natural

companii preocupate de recunoaşterea vorbirii au reuşit să descopere în mod independent valoarea acestei metode şi, de la mijlocul anilor 1980, cea mai mare parte a cercetărilor din domeniul recunoaşterii vorbirii automate s-au bazat pe această abordare. Modelele Markov Ascunse sunt folosite, de asemenea, în sinteza vorbirii – reține, cititorule, că ierarhia neocorticală biologică este utilizată nu numai pentru recunoaşterea input-ului, ci şi pentru producerea output-ului, de exemplu vorbirea şi mişcarea fizică.

Modelele Markov Ascunse Ierarhice, MMAI, sunt folosite şi în sisteme care înțeleg semnificația propozițiilor în limbaj natural, ceea ce înseamnă urcarea în ierarhia conceptuală.

Pentru a înțelege cum funcționează Modelele Markov Ascunse Ierarhice, MMAI, pornim de la o rețea care constă în toate tranzițiile de stare posibile. Metoda cuantificării vectoriale descrisă mai înainte este fundamentală aici, deoarece altminteri nu ar exista prea multe posibilități de luat în considerare.

Iată o topologie inițială simplificată:



O topologie simplă pentru Modelul Markov Ascuns, pentru recunoaşterea a două cuvinte rostite

Mostrele de rostire sunt procesate pe rând, câte una. Pentru fiecare, modificăm în mod recurent probabilitățile tranzițiilor, pentru a reflecta mai bine mostra de input pe care tocmai am procesat-o. Modelele Markov folosite în recunoaşterea vorbirii codifică posibilitatea ca în fiecare fonem să se găsească anumite modele de sunete, modul în care se influențează reciproc fonemele şi ordinea posibilă a fonemelor. Sistemul mai poate să cuprindă şi rețele de probabilitate la niveluri superioare de structură a limbajului, cum ar fi ordinea cuvintelor, includerea de expresii şi aşa mai departe, urcând în ierarhia limbajului.

În vreme ce sistemele de recunoaştere anterioare încorporau reguli specifice referitoare la structurile fonemelor şi succesiuni explicit codificate de lingviştii umani, noului sistem bazat pe Modelele Markov Ascunse Ierarhice nu i s-a spus că în limba engleză există patruzeci şi patru de foneme, nu i s-au dezvăluit succesiunile de vectori probabile pentru fiecare fonem şi nici ce succesiuni de foneme sunt mai probabile decât altele. Am lăsat sistemul să descopere aceste „reguli" singur, din miile de ore de date de limbaj uman care au fost transcrise. Avantajul acestei abordări în raport cu regulile codificate manual este că modelele pun la punct reguli probabilistice de care experții umani sunt rareori conştienți. Am observat că multe dintre regulile pe care sistemul le-a învățat în mod automat din date se deosebeau în moduri subtile, dar importante, de regulile stabilite de experții umani.

Din momentul în care rețeaua a fost pregătită, am încercat să recunoaştem limbajul luând în considerare căi alternative în rețea şi alegând calea cea mai probabilă, dată fiind succesiunea respectivă de vectori de input pe care o văzuserăm. Cu alte cuvinte, dacă vedeam o succesiune de stări care ar fi fost posibil să producă rostirea unui cuvânt, concluzionam că rostirea s-a produs în urma acelei succesiuni corticale. Această simulare de neocortex bazată pe Modelele Markov Ascunse Ierarhice conținea cuvinte-etichetă, astfel încât era capabilă să propună o transcriere a ceea ce auzea.

Am putut pe urmă să ne perfecționăm rezultatele şi mai mult, continuând să antrenăm rețeaua în timp ce o foloseam pentru recunoaştere. După cum am discutat, recunoaşterea şi învățarea concomitente se petrec şi la toate nivelurile ierarhiei noastre neocorticale biologice.

Algoritme evoluționiste (genetice)

Mai există un considerent foarte important: Cum setăm numeroşii parametri care controlează funcționarea unui sistem de recunoaştere a

modelelor? Acestea pot să cuprindă numărul de vectori pe care îl permitem în pasul cuantificării vectoriale, topologia inițială a stărilor ierarhice (înainte ca faza de antrenare a procesului Modelului Markov Ascuns să le reducă), pragul de recunoaștere de la fiecare nivel al ierarhiei, parametrii care controlează manevrarea parametrilor de mărime și multe altele. Putem stabili toate acestea pe baza intuiției, dar rezultatele vor fi departe de a fi optime.

Numim acești parametri „Parametri-Dumnezeu" deoarece sunt stabiliți înaintea metodei de determinare a topologiei Modelelor Markov Ascunse (sau, în cazul biologic, înainte ca persoana să-și învețe lecțiile prin crearea similară a conexiunilor în ierarhia sa corticală). Probabil că termenul este impropriu, dat fiind faptul că aceste detalii inițiale de concepție bazate pe ADN sunt determinate de evoluția biologică, deși unii pot să vadă în acest proces intervenția mâinii lui Dumnezeu (și, cu toate că personal consider că evoluția este într-adevăr un proces spiritual, această discuție își are locul în Capitolul 9).

Când am ajuns la setarea acestor „Parametri ai lui Dumnezeu" în sistemul nostru simulat ierarhic de învățare și recunoaștere, ne-am inspirat din nou de la natură și am decis să-i perfecționăm – în cazul nostru, folosind o simulare a evoluției. Am folosit algoritmii genetici sau evolutivi (AG) care includ reproducerea sexuală și mutațiile.

Iată o descriere simplificată a funcționării acestei metode. Mai întâi, determinăm un mod de codificare a posibilelor soluții pentru o problemă dată. Dacă problema este optimizarea parametrilor de concepție pentru un circuit, atunci definim o listă cu toți parametrii (cu un număr specific de biți atribuit fiecărui parametru) care caracterizează circuitul. Această listă este considerată un cod genetic în algoritmul genetic. Pe urmă, generăm în mod aleatoriu mii de coduri genetice sau mai multe. Fiecare asemenea cod genetic (care reprezintă un set de parametri de concepție) este considerat un organism-„soluție" simulat.

Acum evaluăm fiecare organism simulat într-un mediu simulat, folosind o metodă bine definită de evaluare a fiecărui set de parametri. Evaluarea este o cheie a succesului pentru algoritmul genetic. În exemplul nostru, am rula fiecare program generat de parametri și l-am judeca după criterii adecvate (și-a îndeplinit sarcina, timpul de care a avut nevoie și așa mai departe). Celor mai bune organisme-soluție (celor mai bune concepții) li se permite să supraviețuiască, iar restul sunt eliminate.

Acum îi facem pe toți supraviețuitorii să se înmulțească până când ajung la același număr de creaturi-soluție. Această înmulțire se realizează prin simularea reproducerii sexuale: cu alte cuvinte, creăm noi progenituri care își extrag, fiecare, o parte din codul genetic de la un părinte, iar cealaltă parte de la al doilea părinte. De obicei nu se face nicio distincție între organismele-mascul și cele femelă; este suficient să se genereze o progenitură din oricare doi părinți arbitrari, așadar este vorba aici despre căsătorii între părți de același sex. Probabil că mixajul nu este la fel de interesant ca reproducerea sexuată din lumea naturală, dar aspectul relevant stă în faptul că există doi părinți. Pe măsură ce aceste organisme simulate se înmulțesc, permitem să apară anumite mutații (modificări aleatorii) la nivelul cromozomilor.

Am definit, așadar, o generație de evoluție simulată; acum repetăm aceeași pași pentru fiecare generație ulterioară. La sfârșitul oricărei generații, determinăm măsura în care s-a perfecționat concepția (respectiv, computăm progresul mediu în funcția de evaluare pentru toate organismele care au supraviețuit). Când gradul de progres în evaluarea creaturilor concepute de la o generație la alta devine foarte mic, oprim ciclul recurent și folosim cea (cele) mai bună (bune) concepție (concepții) din ultima generație. (Pentru o descriere a algoritmilor genetici, vezi această notă de final.)[11]

Cheia unui algoritm genetic stă în faptul că creatorii lui umani nu programează o soluție în mod direct; dimpotrivă, lăsăm să apară una printr-un proces recurent de competiție și progres simulate. Evoluția biologică este inteligentă dar lentă, astfel încât, pentru a-i spori inteligența, îi accelerăm foarte mult ritmul greoi. Computerul este suficient de rapid pentru a simula mai multe generații într-un răstimp de câteva ore sau câteva zile; ocazional, le-am lăsat să ruleze săptămâni întregi, pentru a simula sute de mii de generații. Însă trebuie să trecem prin acest proces recurent o singură dată; de îndată ce am lăsat evoluția simulată să-și urmeze cursul, putem aplica regulile evoluate și deosebit de rafinate la problemele reale într-o manieră foarte rapidă. Am folosit sistemele de recunoaștere a vorbirii pentru evoluarea topologiei inițiale a rețelei și a altor parametri fundamentali. Astfel, am folosit două metode de autoorganizare: un algoritm genetic pentru simularea evoluției biologice care a dat naștere unei anumite concepții corticale și Modele Markov Ascunse Ierarhice pentru simularea organizării corticale care însoțește învățarea umană.

O altă cerință majoră pentru succesul unui algoritm genetic rezidă într-o metodă validă de evaluare a fiecărei soluții posibile. Această evaluare trebuie

realizată rapid, deoarece trebuie să țină seamă de multe mii de posibile soluții pentru fiecare generație a evoluției simulate. Algoritmii genetici sunt foarte abili în rezolvarea problemelor cu prea multe variabile, pentru care se cere să se computeze soluții analitice precise. Conceperea unui motor, de exemplu, poate să implice mai bine de o sută de variabile şi impune satisfacerea câtorva zeci de constrângeri; algoritmii genetici folosiți de cercetătorii de la General Electric au putut produce concepte de motoare cu reacție care întrunesc constrângerile mult mai precis decât metodele convenționale.

În cazul folosirii algoritmilor genetici, trebuie, însă, să fii foarte atent la ceea ce anume ceri. Un algoritm genetic a fost folosit pentru rezolvarea problemei depozitării unor blocuri şi a dat o soluție perfectă... cu excepția faptului că presupunea mii de paşi. Programatorii umani au uitat să includă minimalizarea numărului de paşi în funcția lor de evaluare.

Proiectul lui Scott Dave, *Electric Sheep*, este un algoritm genetic care produce lucrări de artă. Funcția de evaluare foloseşte evaluatori umani în cadrul unei colaborări de tip *open-source* în care sunt implicați mii de oameni. Lucrările de artă se modifică în timp şi poți să le vezi la adresa electricsheep.org.

Pentru recunoaşterea vorbirii, combinația dintre algoritmii genetici şi Modelele Markov Ascunse a funcționat extrem de bine. Simularea evoluției cu ajutorul unui algoritm genetic a putut să îmbunătățească substanțial performanța rețelelor pe bază de Modele Markov Ascunse Ierarhice. Ceea ce a adus evoluția a fost net superior concepției noastre inițiale, care se baza pe intuiția noastră.

Pe urmă am experimentat introducerea unei serii de mici variații în sistemul de ansamblu. De exemplu, am produs perturbări (mici modificări aleatorii) ale input-ului. O altă asemenea modificare a constat în a face Modelele Markov alăturate să se „scurgă" unele în altele, determinând rezultatele unui Model Markov să influențeze modelele aflate în „apropiere".

Deşi nu ne-am dat seama la vremea respectivă, genul de ajustări pe care le-am experimentat seamănă foarte mult cu tipurile de modificări care se produc în structurile corticale biologice.

La început, asemenea schimbări afectează performanța (măsurată în acuratețea recunoaşterii). Dar rerulând evoluția (respectiv, rerulând algoritmul genetic) după operarea acestor alterări, sistemul s-a adaptat în consecință, optimizându-se pentru modificările introduse. În general, procedura aceasta restaurează performanța. Când, ulterior, am înlăturat schimbările introduse,

performanța s-a degradat, deoarece sistemul evoluase în direcția compensării schimbărilor. Sistemul adaptat a devenit dependent de schimbări.

Un tip de alterare care a ajutat practic performanța (după rerularea algoritmului genetic) a constat în introducerea unor mici modificări aleatorii în input. Motivul acestei atitudini este binecunoscuta problemă a „overfitting"-ului care apare la sistemele autoorganizate. Există pericolul ca un asemenea sistem să suprageneralizeze în funcție de exemplele specifice conținute în mostra de pregătire. Operând ajustări aleatorii ale input-ului, supraviețuiesc modelele invariabile din date şi prin asta sistemul învață aceste modele mai profunde. Metoda a fost de ajutor numai dacă am rerulat algoritmul genetic cu caracteristica de randomizare activată.

Aceasta introduce o dilemă în înțelegerea circuitelor noastre corticale biologice. S-a observat, de exemplu, că într-adevăr poate exista o slabă scurgere de la o conexiune corticală la alta, rezultată din modul în care se formează conexiunile biologice: se pare că proprietățile electrochimice ale axonilor şi dendritelor sunt supuse efectelor electromagnetice ale conexiunilor din apropiere. Să presupunem că am putea face un experiment prin care să înlăturăm acest efect dintr-un creier real. Acest lucru ar fi dificil de realizat practic, dar nu neapărat imposibil. Să presupunem că am face un asemenea experiment şi am descoperi că circuitele corticale funcționează mai puțin eficient în absența acestei scurgeri neuronale. Am putea concluziona, aşadar, că acest fenomen a reprezentat o intenție foarte inteligentă a evoluției şi a avut o importanță decisivă pentru ca neocortexul să poată ajunge la nivelul său specific de performanță. Am putea scoate în evidență, mai mult, că un asemenea rezultat indică faptul că modelul ordonat al fluxului ascendent de modele în ierarhia conceptuală şi fluxul descendent de predicții în ierarhia conceptuală au fost, de fapt, complicate din cauza influenței complexe dintre conexiuni.

Dar aceasta nu ar fi în mod necesar o concluzie exactă. Să ne gândim la experiența noastră cu un cortex simulat bazat pe Modele Markov Ascunse Ierarhice, în care am implementat o modificare foarte asemănătoare cu încrucişarea interneuronală. Dacă am rerula evoluția cu acest fenomen instituit, performanța ar fi reinstaurată (deoarece procesul evoluției s-a adaptat la el). Dacă am înlocui apoi încrucişarea, performanța ar fi din nou compromisă. În cazul biologic, evoluția (respectiv evoluția biologică) a „rulat" într-adevăr în condițiile instaurării acestui fenomen. Parametri detaliați ai sistemului au fost setați, prin urmare, prin evoluție biologică, pentru a deveni dependenți de aceşti factori, astfel încât modificarea lor va afecta negativ performanța

dacă nu rulăm din nou evoluția. Această abordare este fezabilă în lumea simulată, în care evoluția durează numai câteva zile sau săptămâni, dar în lumea biologică ar necesita zeci de mii de ani.

Prin urmare, cum putem să ne dăm seama dacă o anumită caracteristică de proiectare a neocortexului biologic este sau nu o inovație vitală produsă de evoluția biologică – respectiv una instrumentală pentru nivelul nostru de inteligență – sau pur și simplu un artefact de care proiectarea sistemului este dependentă de-acum, dar fără de care nu ar fi putut să evolueze? Putem răspunde simplu la această întrebare rulând evoluția simulată cu și fără aceste variațiuni particulare ale detaliilor de proiectare (de exemplu, cu și fără conexiuni încrucișate). Putem face acest lucru chiar și în cazul evoluției biologice, dacă examinăm evoluția unei colonii de microorganisme la care generațiile se măsoară în ore, însă nu este practică pentru organisme complexe ca omul. Acesta este un alt mare dezavantaj al biologiei.

Revenind la cercetările noastre din domeniul recunoașterii vorbirii, am descoperit că, dacă rulam evoluția (respectiv un algoritm genetic) separat pe proiectul inițial (1) al Modelelor Markov Ascunse Ierarhice care făceau modelele structurii interne a fonemelor și (2) pe Modelele Markov Ascunse Ierarhice ale structurilor cuvintelor și expresiilor, obțineam rezultate și mai bune. Ambele niveluri ale sistemului foloseau Modele Markov Ascunse Ierarhice, dar algoritmul genetic asigura evoluția unor variațiuni de proiectare între aceste niveluri diferite. Această abordare permitea în continuare realizarea unor modele ale fenomenelor care se petrec între cele două niveluri, cum ar fi întinderea fonemelor care apare adeseori atunci când înșiruim anumite cuvinte unele după altele (de exemplu, „*How are you all doing?*" poate să devină „*How're y'all doing*?").

Este posibil ca un fenomen similar să se fi manifestat în diferite regiuni ale cortexului biologic, prin care au evoluat mici deosebiri bazate pe tipurile de modele de care se ocupau. În vreme ce toate aceste regiuni folosesc același algoritm neocortical fundamental, evoluția biologică a avut suficient timp la dispoziție pentru a face acordul fin al proiectului fiecăreia dintre ele, ca să devină optime pentru modelele lor. Pe de altă parte însă, așa cum am arătat mai înainte, specialiștii în neuroștiință și neurologii au observat în aceste zone o plasticitate substanțială, fapt ce sprijină ideea unui algoritm general neocortical. Dacă metodele fundamentale pentru fiecare regiune ar fi radical diferite, o asemenea interșanjabilitate între regiunile corticale nu ar fi posibilă.

Sistemele pe care le-am creat în cursul cercetărilor noastre folosind această combinație de metode de autoorganizare au fost încununate de un mare succes. În cazul recunoașterii vorbirii, ele au reușit pentru prima oară să facă față vorbirii continue normale și unui vocabular relativ nerestricționat. Am izbutit să atingem o acuratețe foarte înaltă pentru o mare varietate de vorbitori, accente și dialecte. Actualul stadiu de vârf, în timp ce scriu această carte, este reprezentat de un produs numit Dragon Naturally Speaking (Version 11.5) pentru PC de la Nuance (anterior, Kurzweil Computer Products). Le sugerez tuturor să îl încerce dacă sunt sceptici în privința performanțelor sistemelor contemporane de recunoaștere a vorbirii – acuratețea este adeseori de 99 la sută sau chiar mai mare după câteva minute de antrenare pentru vocea respectivului vorbitor, la vocabulare relativ nerestricționate. Dragon Dictation este o aplicație mai simplă, dar totuși impresionantă pentru iPhone, și nu a avut nevoie de antrenare vocală. Siri, asistentul personal al iPhone-urilor Apple contemporane, folosește aceeași tehnologie de recunoaștere a vorbirii, cu extensii menite a-l ajuta să facă față înțelegerii limbii naturale.

Performanța acestor sisteme este o mărturie a forței matematicii. Cu ajutorul lor, computăm practic ceea ce se petrece în neocortexul vorbitorului – chiar dacă nu avem acces direct la creierul persoanei respective – ca pas înainte vital în recunoașterea spuselor unei persoane și, în cazul unor sisteme ca Siri, a înțelesului mesajelor verbale respective. Am putea să ne întrebăm: *Dacă ne-am uita de-adevăratelea în neocortexul vorbitorului, am vedea conexiunile și greutățile corespunzătoare Modelelor Markov Ascunse Ierarhice computate de software?* Aproape sigur, nu am găsi o potrivire precisă; structurile neuronale s-ar deosebi în mod invariabil de modelele din computer sub aspectul mai multor detalii. Cu toate acestea, aș susține în continuare că trebuie să existe o echivalență matematică esențială, la un grad înalt de precizie, între biologie și încercarea noastră de a o imita; altminteri, aceste sisteme n-ar funcționa atât de bine cum o fac.

LISP

LISP (LISt Processor) este un limbaj de computer propus inițial de pionierul inteligenței artificiale John McCarthy (1927-2011), în anul 1958. După cum sugerează și denumirea, LISP se ocupă de liste. Fiecare

instrucțiune LISP este o listă de elemente; fiecare element este o altă listă sau un „atom", care reprezintă un articol ireductibil, constituind fie un număr, fie un simbol. O listă inclusă într-o listă poate fi lista supremă, de unde rezultă că LISP este capabil de recurență. Instrucțiunile LISP pot fi recurente şi dacă lista include o listă şi aşa mai departe, până când este specificată lista originală. Pentru că listele pot să includă liste, LISP este capabil şi de procesare ierarhică. Lista poate fi condițională, astfel încât să „emită" numai dacă elementele ei sunt satisfăcute. În acest fel, ierarhiile de asemenea condiționalități pot fi folosite pentru identificarea calităților din ce în ce mai abstracte ale unui model.

LISP a ajuns un răcnet în comunitatea inteligenței artificiale în anii 1970 şi la începutul anilor '80. Fantezia pasionaților de LISP ai deceniului inițial susținea că limbajul acesta oglindeşte modul de funcționare a creierului uman şi că orice proces inteligent poate fi codificat cel mai uşor şi mai eficient în LISP. A urmat un miniboom în companiile de „inteligență artificială" care ofereau interpreți LISP şi produse LISP asociate, dar, la mijlocul anilor 1980, când a devenit evident că LISP, ca atare, nu era o scurtătură pentru crearea proceselor inteligente, bula investițiilor a explodat.

S-a dovedit însă că pasionații LISP nu greşiseră întru totul. În esență, toți recognizatorii de modele din neocortex pot fi considerați instrucțiuni LISP – fiecare dintre ei constituind o listă de elemente şi fiecare element putând fi o altă listă. Neocortexul este, prin urmare, angajat în procesarea listelor de o natură simbolică, similar cu ceea ce se petrece într-un program LISP. Mai mult de-atât, procesează toate cele 300 de milioane de „instrucțiuni" gen LISP simultan.

Pe de altă parte, din lumea LISP lipseau două caracteristici importante, dintre care una era învățarea. Trebuia ca programele LISP să fie codificate linie cu linie de programatori umani. Au existat tentative de codificare automată a acestor programe, folosindu-se o varietate de metode, dar acestea nu au făcut parte integrantă din conceptul limbajului. Neocortexul, în schimb, se programează pe el însuşi, umplându-şi „declarațiile" (respectiv listele) cu informații semnificative şi executabile din propria experiență şi din propriile bucle de feedback. Acesta este un principiu-cheie al modului de funcționare a neocortexului: fiecare dintre recognizatorii săi de modele (respectiv fiecare instrucțiune LISP) este capabil(ă) să îşi realizeze propria listă şi să se conecteze atât în sus, cât şi în jos la alte liste. A doua diferență rezidă în mărimea parametrilor. Se poate crea o variantă de LISP (codificată în LISP) care să

permită manevrarea unor asemenea parametri, dar această variantă nu face parte din limbajul de bază.

LISP este consecvent filosofiei originale a domeniului inteligenței artificiale, care presupune găsirea de soluții inteligente la probleme şi codificarea acestor soluții direct în limbajul-maşină. Prima tentativă de creare a unei metode de autoorganizare care să învețe din propria experiență – rețelele neuronale – n-a avut succes, deoarece n-a asigurat un mijloc pentru modificarea topologiei sistemului ca răspuns la învățare. Modelele Markov Ascunse Ierarhice au asigurat asta în mod eficient prin mecanismul lor de tăiere. Astăzi, Modelele Markov Ascunse Ierarhice împreună cu verii lor matematici constituie elemente majore în lumea inteligenței artificiale.

Un corolar al observării similitudinilor dintre LISP şi structura de listă a neocortexului reprezintă un argument adus de cei ce insistă că, totuşi, creierul este prea complicat pentru ca noi să-l putem înțelege. Criticii scot în evidență faptul că un creier are miliarde de conexiuni şi, de vreme ce fiecare dintre ele trebuie să se afle acolo potrivit proiectului, ele constituie echivalentul a mii de miliarde de linii de cod. Aşa cum am văzut, am estimat că la nivelul neocortexului există ceva de ordinul a 300 de milioane de procesori de modele – sau 300 de milioane de liste, cu proprietatea că fiecare element dintr-o listă indică o altă listă (sau, la nivelul conceptual cel mai redus, un model de bază ireductibil din afara neocortexului). Dar cifra de 300 de milioane este una rezonabil de mare de instrucțiuni LISP şi e într-adevăr mai mare decât orice program scris de om şi aflat în funcțiune.

În orice caz, trebuie să reținem că aceste liste nu sunt practic specificate în concepția inițială a sistemului nervos. Creierul le creează din propriile sale experiențe. Acesta este secretul-cheie al neocortexului. Procesele care realizează această autoorganizare sunt mult mai simple decât cele 300 de milioane de instrucțiuni care constituie capacitatea neocortexului. Aceste procese sunt specificate în genom. Aşa cum voi demonstra în Capitolul 11, volumul de informație unică din genom (după comprimarea fără pierdere) aplicată creierului este de aproximativ 25 de milioane de biți – echivalentul a mai puțin de un milion de rânduri de cod. Complexitatea algoritmică propriu-zisă este chiar mai mică de-atât, pentru că, din cele 25 de milioane de biți de informație genetică, partea cea mai însemnată se referă la nevoile biologice ale neuronilor şi nu neapărat la capacitatea lor de procesare a informației. Însă chiar şi 25 de milioane de biți de informație de concepție reprezintă un nivel de complexitate căruia putem să-i facem față.

Sisteme ierarhice de memorie

După cum am spus în Capitolul 3, Jeff Hawkins şi Dileep George au pus la punct în anii 2003 şi 2004 un model de neocortex care încorporează listele ierarhice descrise în cartea din 2004 a lui Hawkins şi Blakeslee *On Intelligence – Despre inteligență*. O prezentare mai actualizată şi foarte elegantă a memoriei temporale ierarhice poate fi găsită în dizertația de doctorat din anul 2008 a lui Dileep George.[12] Numenta a implementat-o într-un sistem numit NuPIC (Numenta Platform for Intelligent Computing – Platforma Numenta pentru computație inteligentă) şi a pus la punct sisteme de recunoaştere a modelelor şi de căutare de date pentru clienți cum ar fi Forbes şi Power Analytics Corporation. După ce a lucrat la Numenta, George a fondat o nouă companie, intitulată Vicarious Systems, cu finanțare de la Founder Fund (condusă de Peter Thiel, investitorul care a stat în spatele Facebook, şi de Sean Parker, primul preşedinte Facebook) şi de la Good Ventures, condusă de Dustin Maskovitz, cofondator Facebook. George a comunicat progrese semnificative în modelarea, învățarea şi recunoaşterea automată a informațiilor cu un număr substanțial de ierarhii. Şi-a numit sistemul „rețea corticală recursivă" şi intenționează să realizeze aplicații pentru imagistica şi robotica medicală, dar nu numai. Tehnica Modelelor Markov Ascunse Ierarhice este foarte asemănătoare din punct de vedere matematic cu aceste sisteme de memorie ierarhică, mai ales dacă permitem sistemului bazat pe Modele Markov Ascunse Ierarhice să îşi organizeze propriile conexiuni între modulele de recunoaştere a modelelor. După cum am menționat anterior, Modelele Markov Ascunse Ierarhice mai asigură un element adițional important, care este modelarea distribuției aşteptate a magnitudinii (într-un continuum oarecare) pentru fiecare input la computarea probabilității existenței modelului avut în vedere. Recent, am lansat o nouă companie, denumită Patterns, Inc., care intenționează să dezvolte modele neocorticale autoorganizate ierarhic pe baza Modelelor Markov Ascunse Ierarhice şi tehnici asociate în scopul înțelegerii limbajului natural. Un accent important va fi pus pe abilitatea sistemului de a-şi concepe propriile ierarhii într-o manieră similară neocortexului biologic. Sistemul avut în vedere de noi va citi în mod continuu o vastă gamă de materiale, cum ar fi Wikipedia şi alte resurse de cunoştințe, ascultând de asemenea tot ce i se spune şi urmărind tot ce se scrie (dacă i se permite). Scopul este să devină un prieten util,

care să-ți răspundă la întrebări – chiar dinainte să le formulezi – și să-ți dea informații și ponturi utile în viața de zi cu zi.

Frontiera în mișcare a inteligenței artificiale: ascensiunea în ierarhia competenței

1. Un discurs lung și plictisitor ținut de un *topping* spumos de plăcintă.
2. Articol de îmbrăcăminte purtat de un copil, eventual la bordul unui vas de operă.
3. Căutat de doisprezece ani pentru crima de a-i fi mâncat pe războinicii regelui Hrothgar; cazul îi este dat ofițerului Beowulf.
4. Poate să însemne să dezvolți treptat în minte sau să porți în timpul sarcinii.
5. Ziua Națională a Învățătorului și Ziua Derby-ului din Kentucky.
6. Wordsworth a spus că ei se avântă, dar niciodată nu hoinăresc.
7. Cuvânt din patru litere pentru fierul care se pune pe copita calului sau pe masa de joc dintr-un cazinou.
8. În actul trei al operei lui Verdi din anul 1846, acest Bici al lui Dumnezeu este înjunghiat și omorât de iubita lui, Odabella.

— Exemple de întrebări *Jeopardy!*,
la toate Watson dând răspunsuri corecte.
Răspunsurile sunt: meringue harangue
[joc de cuvinte intraductibil]; șorț; Grendel;
gestație; mai; ciocârlie; potcoavă.
La a treia întrebare, Watson a răspuns:
„Ce este Attila?" Gazda a replicat: „Fii mai exact".
Watson a clarificat cu „Ce este Attila Hunul?",
ceea ce este corect.

Tehnicile computerului pentru dezlegarea indiciilor la *Jeopardy!* seamănă cu ale mele. Mașina se concentrează pe cuvintele-cheie din indiciu, apoi își consultă memoria (în cazul lui Watson, o bancă de date de 15 terabiți de cunoștințe umane) în căutarea unor aglomerări de asociații cu aceste cuvinte. Confruntă cu rigurozitate principalele răspunsuri cu informațiile contextuale pe care le poate găsi: denumirea categoriei; genul de răspuns care e așteptat; momentul, locul și genul masculin sau feminin

sugerate de indiciu și așa mai departe. Iar când se simte suficient de „sigur", decide să apese soneria. Tot acest proces este instantaneu, intuitiv pentru un jucător uman la *Jeopardy!*, dar eu am fost convins că, sub glugă, și creierul meu făcea mai mult sau mai puțin același lucru.

— Ken Jennings, campion uman la *Jeopardy!* care a pierdut în fața lui Watson

Eu, unul, îi întâmpin cu brațele deschise pe noii noștri stăpâni roboți.

— Ken Jennings, parafrazându-i pe Simpsoni, după ce a pierdut în fața lui Watson

Dumnezeule mare! [Watson] este mai inteligent decât jucătorul mediu de *Jeopardy!* în ceea ce privește răspunsurile la întrebările de la *Jeopardy!*. Asta înseamnă impresionant de inteligent.

— Sebastian Thrun, fost director al Laboratorului de Inteligență Artificială de la Stanford

Watson nu înțelege nimic. E doar un buldozer mai mare.

— Noam Chomsky

Inteligența artificială există pretutindeni în jurul nostru – nu mai avem niciun cuvânt de spus în privința asta. Simplul gest al conectării cu cineva prin intermediul unui mesaj de tip text, e-mail sau apel pe un telefon mobil folosește algoritmi inteligenți pentru rutarea informației. Aproape toate produsele pe care le atingem sunt proiectate inițial printr-o colaborare între o inteligență artificială și una umană, iar apoi sunt realizate în fabrici automatizate. Dacă toate sistemele de inteligență artificială ar decide ca mâine să intre în grevă, civilizația noastră ar fi paralizată: nu am mai putea scoate bani de la bancă și până și banii ar dispărea; comunicațiile, transportul și producția s-ar opri brusc. Din fericire, mașinile noastre inteligente nu sunt deocamdată atât de inteligente încât să organizeze o asemenea conspirație.

Nouă astăzi în inteligența artificială este natura visceral impresionantă a exemplelor disponibile în spațiul public. Bunăoară, să ne gândim la automobilele Google care se conduc singure (în momentul în care scriu aceste rânduri, ele au trecut de 320.000 de kilometri parcurși în orașe mari și mici), o tehnologie ce va conduce la un număr semnificativ redus de accidente rutiere, la creșterea capacității drumurilor, la eliminarea necesității ca omul să

îşi asume povara şofatului şi la multe alte beneficii. Deja, automobilele fără şofer sunt, practic, legale pe drumurile publice din Nevada, deşi sunt supuse unor restricţii, dar utilizarea lor de către publicul larg din întreaga lume nu este aşteptată înaintea următorului deceniu. Tehnologia care urmăreşte drumul în mod inteligent şi îl avertizează pe şofer cu privire la pericolele iminente este deja instalată pe automobile. O asemenea tehnologie se bazează parţial pe modelul de succes al procesării vizuale la nivelul creierului, creată de Tomaso Poggio de la MIT. Denumită MobilEye, ea a fost pusă la punct de Amnon Shashuna, fost doctorand al lui Poggio. Tehnologia este capabilă să îl avertizeze pe şofer cu privire la pericole cum ar fi o coliziune iminentă sau un copil care aleargă în calea automobilului şi a fost utilizată recent de producători precum Volvo şi BMW.

În această secţiune a cărţii mă voi concentra asupra tehnologiilor limbajului – şi asta din mai multe motive. Deloc surprinzător, natura ierarhică a limbajului oglindeşte fidel natura ierarhică a gândirii noastre. Limbajul vorbit a fost prima noastră tehnologie, iar limbajul scris – cea de-a doua. Propriile mele cercetări în domeniul inteligenţei artificiale, după cum a demonstrat acest capitol, s-au concentrat puternic pe limbaj. În cele din urmă, stăpânirea limbajului este o capacitate cu o extraordinară forţă de influenţă. Watson a citit deja sute de milioane de pagini pe web şi stăpâneşte toate cunoştinţele conţinute în aceste documente. În cele din urmă, maşinile vor fi capabile să acumuleze întreaga cunoaştere de pe web – care, în esenţă, reprezintă toată cunoaşterea civilizaţiei om-maşină.

Matematicianul englez Alan Turing (1912–1954) şi-a întemeiat testul eponim pe capacitatea computerului de a conversa în limbaj natural folosind mesaje text.[13] Turing a considerat că întreaga inteligenţă umană a fost întrupată şi reprezentată de limbaj şi că nicio maşină nu poate trece Testul Turing prin simple trucuri de limbaj. Deşi Testul Turing este un joc care implică limbajul scris, Turing credea că singurul mod în care poate să-l treacă un computer este ca maşina să posede într-adevăr echivalentul inteligenţei de nivel uman. Criticii au avansat ipoteza că un adevărat test al inteligenţei de nivel uman ar trebui să includă şi stăpânirea informaţiei vizuale şi auditive.[14] Deoarece multe dintre propriile mele proiecte ce ţin de inteligenţa artificială presupun învăţarea computerului să stăpânească informaţii senzoriale cum ar fi vorbirea umană, formele literelor şi sunetele muzicale, s-ar aştepta din partea mea să fiu partizanul includerii acestor tipuri de informaţie într-un test adevărat de inteligenţă. Totuşi, sunt de acord cu intuiţia iniţială a lui Turing că

versiunea exclusiv text a Testului Turing este suficientă. Adăugarea input-ului vizual sau auditiv nu ar face cu adevărat testul mai greu de trecut.

Nu trebuie să fii expert în inteligență artificială ca să fii impresionat de performanța lui Watson la *Jeopardy!* Cu toate că înțeleg la un nivel rezonabil metodologia folosită la multe dintre subsistemele sale cheie, asta nu-mi diminuează reacția emoțională atunci când îl văd în acțiune. Nici măcar o înțelegere perfectă a modului de funcționare a tuturor sistemelor lui componente – la care nu a ajuns încă nimeni – nu te-ar putea ajuta să anticipezi felul cum va reacționa de fapt Watson la o situație dată. El conține sute de subsisteme care interacționează și fiecare dintre acestea analizează milioane de ipoteze concurente în același timp, deci anticiparea rezultatului este imposibilă. O analiză temeinică – *post-factum* – a deliberărilor lui Watson pentru o singură căutare de trei secunde ar necesita secole de viață de om.

Ca să-mi continuu propria istorie, la sfârșitul anilor 1980 și începutul anilor '90, am început să lucrăm la înțelegerea limbajului natural în câteva domenii limitate. Poți să-i vorbești unuia dintre produsele noastre, denumit Kurzweil Voice, despre orice, câtă vreme subiectul are legătură cu editarea documentelor. (De exemplu, „Mută al treilea paragraf de la pagina anterioară aici.") Produsul a dat rezultate destul de bune în acest domeniu limitat dar util. Am creat de asemenea sisteme cu cunoștințe din domeniul medical, astfel încât medicii să-și poată dicta rapoartele despre pacienți. Sistemul nostru avea cunoștințe suficiente din domenii cum ar fi radiologia și patologia ca să-i poată adresa întrebări medicului dacă ceva din raport i se părea neclar și putea să-l călăuzească pe acesta în procesul de raportare. Sistemele de raportare medicală au evoluat, devenind o afacere de miliarde de dolari la Nuance.

Înțelegerea limbajului natural, în special ca extensie a recunoașterii automate a vorbirii, a devenit un fapt cotidian. În momentul în care scriu această carte, Siri, asistentul personal automat de pe iPhone 4S, produce furori în lumea computației mobile. Îi poți cere lui Siri să facă orice e capabil să facă un smartphone care se respectă (de exemplu: „Unde pot să găsesc aici, primprejur, mâncare indiană?" sau „Transmite-i un mesaj text soției că sunt pe drum" sau „Ce părere are lumea de noul film al lui Brad Pitt?") și, de cele mai multe ori, Siri se va conforma. Siri poate susține și o conversație redusă nonproductivă. Dacă întrebi care este sensul vieții, îți va răspunde „42", replică în care pasionații *The Hitchhiker's Guide to the Galaxy – Ghidul autostopistului în galaxie* vor recunoaște „răspunsul la întrebarea supremă

despre viață, Univers, tot". La întrebările care țin de cunoaştere (inclusiv cea despre sensul vieții) răspunde Wolfram Alphy, descris la pagina 164. Există o întreagă lume de „chatboți" care nu fac nimic altceva în afară de a se angaja în conversații mărunte. Dacă vrei să vorbeşti cu chatbotul nostru pe nume Ramona, du-te la website-ul KurzweilAI.net şi dă clic pe „Chat cu Ramona".

Unii oameni mi s-au plâns în legătură cu incapacitatea lui Siri de a răspunde la anumite comenzi, dar îmi amintesc de multe ori că este vorba despre aceiaşi oameni care se plâng consecvent şi de oamenii care asigură servicii. Uneori sugerez să încercăm împreună şi adeseori funcționează mai bine decât se aşteaptă aceşti oameni. Plângerile îmi amintesc de povestea cu câinele care joacă şah. Unui neîncrezător care îl întreabă, stăpânul câinelui îi răspunde: „Da, e adevărat, chiar joacă şah, dar are finalurile slabe". Încep să apară concurenți eficienți, cum ar fi Google Voice Search.

Faptul că publicul face acum conversație în limbaj natural vorbit cu computerele ținute în mână marchează o nouă eră. Este ceva obişnuit ca oamenii să respingă semnificația unei tehnologii de primă generație din cauza limitărilor ei. După vreo câțiva ani, când tehnologia funcționează bine, oamenii încă îi mai resping importanța pentru că, asta este, nu mai e nouă. Acestea fiind spuse, Siri funcționează impresionant pentru un produs de primă generație şi este limpede că această categorie de produse se va îmbunătăți din ce în ce.

Siri foloseşte tehnologiile de recunoaştere a vorbirii bazate pe Modele Markov Ierarhice de la Nuance. Extensiile în limbaj natural au fost puse la punct prima oară de proiectul intitulat „CALO", finanțat de DARPA.[15] Siri a fost îmbunătățit cu tehnologiile pentru limbajul natural ale Nuance, care oferă şi o tehnologie foarte asemănătoare intitulată Dragon Go![16]

Metodele folosite pentru înțelegerea limbii naturale sunt similare cu Modelele Markov Ascunse şi foarte des sunt folosite Modelele Markov Ascunse Ierarhice. În vreme ce unele dintre aceste sisteme nu sunt etichetate precis ca folosind Modele Markov Ascunse sau Modele Markov Ascunse Ierarhice, suportul matematic este practic identic. Toate implică ierarhii de succesiuni liniare în care fiecare element are o greutate, conexiuni autoadaptative şi un sistem de ansamblu autoorganizat, bazat pe învățarea datelor. De obicei, învățarea continuă în cursul utilizării propriu-zise a sistemului. Această abordare corespunde structurii ierarhice a limbii naturale – este doar o extensie naturală a scării conceptuale de la părți de vorbire, la cuvinte, expresii şi structuri semantice. Ar fi logic să se ruleze un algoritm genetic

pe aceşti parametri care controlează exact algoritmul de învățare specific acestei clase de sisteme de învățare ierarhice şi care determină detaliile optime ale algoritmului.

Pe parcursul ultimului deceniu, s-a înregistrat o modificare a felului în care sunt create structurile ierarhice. În anul 1984, Douglas Lenat (născut în 1950) a început lucrul la ambițiosul Proiect Cyc (de la enCYClopedic), care urmărea să creeze reguli capabile să codifice cunoştințele cotidiene „de bun-simț". Aceste reguli erau organizate într-o colosală ierarhie şi fiecare implica – din nou – o succesiune liniară de stări. De exemplu, o regulă Cyc poate spune că un câine are un chip. Pe urmă, Cyc se poate conecta la regulile generale privitoare la structura chipurilor: un chip are doi ochi, un nas, o gură şi aşa mai departe. Nu trebuie să avem un set de reguli pentru chipul unui câine şi altul pentru chipul unei pisici, cu toate că, desigur, ne-am putea dori să includem reguli suplimentare privitoare la felul în care se deosebeşte chipul câinelui de cel al pisicii. Sistemul include, de asemenea, un motor de interferență: dacă avem reguli care afirmă că un cocker spaniel este câine, câinii sunt animale, iar animalele se hrănesc şi dacă am întreba motorul de interferență dacă cocker spanielii mănâncă, sistemul ar răspunde că da, cocker spanielii mănâncă. Pe parcursul următorilor douăzeci de ani şi după mii de ani-persoană de eforturi, au fost scrise şi testate peste un milion de asemenea reguli. Fapt interesant, limbajul pentru scrierea regulilor Cyc – denumit CycL – este aproape identic cu LISP.

Între timp, o şcoală de gândire opusă considera că abordarea cea mai bună pentru înțelegerea limbii naturale şi pentru crearea sistemelor inteligente în general se bazează pe învățarea automată prin contactul zilnic cu un număr foarte mare de ipostaze ale fenomenului pe care încerca să-l stăpânească sistemul. Un exemplu elocvent al unui asemenea sistem este Google Translate, care poate să traducă în şi din cincizeci de limbi. Asta înseamnă 2.500 de direcții de traducere diferite, deşi, pentru majoritatea perechilor de limbi, în loc să se facă traducerea directă a limbii 1 în limba 2, sistemul traduce limba 1 în engleză şi apoi engleza în limba 2. Acest fapt reduce numărul de translatoare Google care trebuie construite la nouăzeci şi opt (plus un număr de perechi nonengleză pentru care există traducere directă). Translatoarele Google nu folosesc reguli gramaticale; dimpotrivă, ele creează vaste baze de date pentru fiecare pereche de limbi cu traduceri frecvente pe baza unui corpus mare în genul „Piatra Rosetta" de documente traduse între aceste două limbi. Pentru cele şase limbi care constituie limbile

oficiale ale Națiunilor Unite, Google a utilizat documente ale Națiunilor Unite, așa cum sunt publicate acestea în fiecare dintre cele șase limbi. În cazul limbilor mai puțin obișnuite, au fost folosite alte idiomuri. Rezultatele sunt adeseori impresionante. DARPA organizează concursuri anuale pentru cele mai bune sisteme de traducere automată în și din diferite perechi de limbi, iar Google Translate câștigă adeseori pentru anumite perechi, depășind sisteme create direct de lingviști umani.

Pe parcursul ultimului deceniu, două intuiții majore au influențat profund domeniul înțelegerii limbilor naturale. Prima este legată de ierarhii. Cu toate că abordarea Google a debutat cu asocierea succesiunilor plate de cuvinte dintr-o limbă în alta, natura inerent ierarhică a limbii s-a insinuat în mod inevitabil în această operație. Sisteme care încorporează metodic învățarea ierarhică (așa cum sunt Modelele Markov Ascunse Ierarhice) au asigurat o performanță semnificativ mai bună. Pe de altă parte însă, asemenea sisteme nu se construiesc chiar automat. Oamenii trebuie să învețe aproximativ o ierarhie conceptuală o dată și același lucru este valabil și pentru sistemele computerizate, astfel încât procesul de învățare trebuie gestionat cu grijă.

Cealaltă intuiție este că regulile manuale funcționează foarte bine în cazul unui miez de cunoștințe comune de bază. Pentru traducerea pasajelor scurte, această abordare dă adeseori rezultate precise. De exemplu, DARPA a clasificat translatoarele pe bază de reguli din chineză în engleză mai bine decât Google Translate pentru pasaje scurte. Pentru ceea ce se numește trena limbii, care se referă la milioane de expresii ocazionale și de concepte folosite în ele, acuratețea sistemelor bazate pe reguli se apropie de o asimptotă inacceptabil de coborâtă. Dacă plotăm acuratețea înțelegerii limbii naturale raportată la volumul de date de pregătire utilizate, constatăm că sistemele bazate pe date înregistrează inițial performanțe superioare, dar scad la o acurateță medie de circa 70 la sută. Într-un contrast acut, sistemele statistice pot ajunge până la 90 la sută în privința acurateței, dar necesită un volum foarte mare de date pentru atingerea acestui nivel.

Adeseori avem nevoie pentru pregătire de o combinație formată dintr-o performanță cel puțin moderată la un nivel redus de date și de oportunitatea de a ajunge la o acurateță ridicată cu un volum mai mare. Atingerea rapidă a unei performanțe moderate ne permite să scoatem pe piață un sistem, ulterior acesta adunând automat date de pregătire, pe măsură ce este folosit de oameni. În acest fel, se poate atinge un nivel foarte ridicat de învățare în intervalul de timp în care este folosit sistemul, iar acuratețea acestuia se

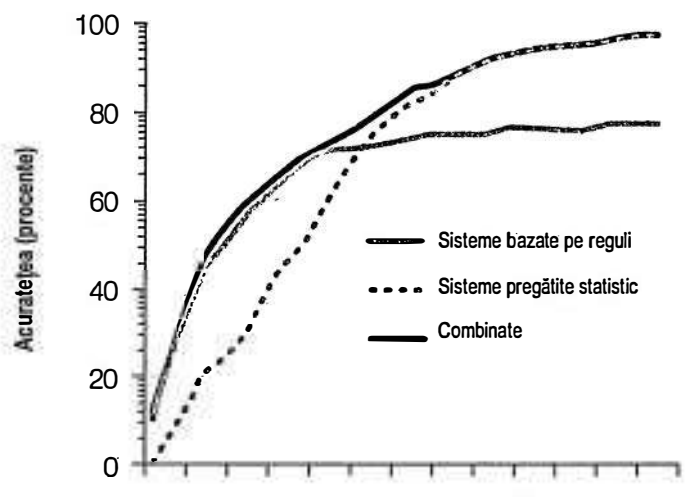
îmbunătățește. Învățarea statistică trebuie să fie pe deplin ierarhică pentru a reflecta natura limbii, care oglindește, la rându-i, modul de funcționare a creierului uman.

Așa funcționează Siri și Dragon Go! – folosind reguli pentru fenomenele cele mai obișnuite și sigure și învățând apoi „trena" limbii atunci când ajung în mâinile utilizatorilor reali. Când și-a dat seama că a atins un plafon de performanță bazat pe reguli codificate manual, echipa Cyc a adoptat și ea această abordare. Regulile codificate manual asigură două funcții esențiale. Prima: oferă o acuratețe inițială adecvată, astfel încât un sistem în faza de încercări poate fi setat pentru utilizare largă, situație în care se va îmbunătăți automat. A doua: asigură o bază solidă pentru nivelurile inferioare ale ierarhiei conceptuale, astfel încât învățarea automată poate să înceapă de la niveluri conceptuale mai înalte.

După cum am menționat anterior, Watson constituie un exemplu impresionant de combinare a regulilor codificate manual cu învățarea statistică ierarhică. IBM a mixat mai multe programe de prim rang de limbă naturală pentru crearea unui sistem care să joace *Jeopardy!* în limbaj natural! În intervalul 14-16 februarie 2011, Watson a concurat cu doi dintre jucătorii umani de vârf: Brad Rutter, care a câștigat mai mulți bani decât oricine altcineva la acest concurs de televiziune cu întrebări și răspunsuri și Ken Jennings, care anterior participase la campionatul de *Jeopardy!* pe durata record de șaptezeci și cinci de zile.

În cadrul contextului, în prima mea carte, *The Age of Intelligent Machines* (*Epoca mașinilor inteligente*, Editura Paralela 45), scrisă la mijlocul anilor 1980, am anticipat că, înainte de 1998, un computer va câștiga campionatul mondial de șah. Am mai anticipat (1) că, atunci când se va întâmpla acest lucru, fie ne vom retrograda opinia cu privire la inteligența umană, fie ne-o vom îmbunătăți pe cea privitoare la inteligența mașinii, fie vom minimaliza importanța șahului și (2) că, dacă ne va fi călăuză istoria, vom minimaliza șahul. Ambele lucruri s-au întâmplat în anul 1997. Când supercomputerul jucător de șah produs de IBM și denumit Deep Blue l-a înfrânt pe campionul mondial la șah Garry Kasparov, am fost întâmpinați imediat cu argumentul că era de așteptat ca un computer să câștige, deoarece computerele sunt mașini logice, iar șahul, în ultimă instanță, este un joc de logică. Astfel, victoria lui Deep Blue a primit sentința că nu este nici surprinzătoare și nici semnificativă. Mulți dintre criticii mașinii au mers mai departe și au susținut că un computer nu va putea stăpâni niciodată

subtilitățile limbii umane, inclusiv metaforele, sinonimele, jocurile de cuvinte, formulările cu dublu înțeles şi umorul.



Acuratețea sistemelor de înțelegere a limbii naturale vorbite, exprimată în funcție de volumul de date pentru pregătire. Cea mai bună abordare este să se combine regulile pentru „miezul" limbii cu o abordare bazată pe date pentru „trena" limbii

Acesta este cel puțin un motiv pentru care Watson reprezintă un reper atât de important: *Jeopardy!* este o sarcină sofisticată şi plină de provocări legate de limbă. În mod curent, întrebările *Jeopardy!* includ multe dintre fanteziile limbii umane. Ceea ce poate nu devine atât de evident pentru mulți observatori este că, pe lângă obligația lui Watson de a stăpâni limba în contextul întrebărilor neaşteptate şi întortocheate, în ecuație a intervenit şi faptul că, în marea lor parte, cunoştințele lui nu au fost codificate manual. Watson deținea aceste cunoştințe după ce citise practic 200 de milioane de pagini de documente în limbă naturală, inclusiv toată *Wikipedia* şi alte enciclopedii, acumulând 4 mii de miliarde de biți de cunoştințe bazate pe limbă. După cum bine ştie cititorul acestei cărți, *Wikipedia* nu e scrisă în LISP sau CycL, ci în propoziții naturale, care conțin toate ambiguitățile şi complexitatea inerente limbii. Watson a trebuit să consulte toate cele 4 mii de miliarde de caractere din materialul lui de referință pentru a răspunde la o întrebare. (Îmi dau

seama că „întrebările" de la *Jeopardy!* sunt de fapt răspunsuri care își caută întrebările, dar acesta este un aspect tehnic – și, în ultimă instanță, ele sunt totuși întrebări.) Dacă Watson poate înțelege și răspunde la întrebări în trei secunde pe baza informațiilor cuprinse în 200 de milioane de pagini, nimic nu poate opri sisteme similare să citească celelalte miliarde de documente de pe web. Și, într-adevăr, această operațiune este în curs de desfășurare chiar acum [în momentul scrierii cărții – n.trad.].

În perioada în care puneam la punct sistemele de recunoaștere a caracterelor și vorbirii și sistemele-pionier de înțelegere a limbii naturale, din anii 1970 și până în anii 1990, am folosit o metodologie de încorporare a unui „manager-expert". Puneam la punct mai multe sisteme care făceau același lucru, dar în fiecare dintre ele încorporam mici abordări diferite. Unele diferențe erau subtile, ca de pildă variații ale parametrilor care controlează latura matematică a algoritmului învățării. Alte diferențe erau fundamentale, ca de exemplu includerea sistemelor bazate pe reguli în locul sistemelor de învățare ierarhic-statistică. Managerul-expert era în sine un program software conceput pentru a învăța punctele-forte și punctele slabe ale diferitelor sisteme prin examinarea performanțelor acestora în situații ținând de lumea reală. El se baza pe ideea că aceste puncte-forte erau ortogonale; respectiv, un sistem tindea să fie forte acolo unde altul tindea să fie slab. Și, într-adevăr, performanța însumată a unor asemenea sisteme combinate cu un manager-expert aflat la conducere a fost cu mult mai mare decât performanțele oricăruia dintre sistemele individuale.

Watson funcționează la fel. Folosind o arhitectură denumită UIMA (Unstructured Information Management Architecture – Arhitectură nestructurată de gestionare a informației), Watson conține literalmente sute de sisteme diferite – multe dintre componentele sale de limbi individuale sunt identice cu cele folosite pentru sistemele de înțelegere a limbii naturale disponibile în mod public – toate încercând fie să dea un răspuns direct la o întrebare de *Jeopardy!,* fie, cel puțin, să realizeze o anumită clarificare a întrebării. UIMA funcționează, în esență, ca un manager-expert pentru combinarea inteligentă a rezultatelor sistemelor independente și depășește cu mult sistemele anterioare – de felul celui pe care l-am pus noi la punct la compania predecesoare Nuance – prin faptul că sistemele lui individuale pot contribui la rezultat fără să dea în mod necesar răspunsul final. Este suficient ca un subsistem să ajute la îngustarea domeniului soluției. De asemenea, UIMA e capabilă să computeze măsura în care are încredere în răspunsul final. Și creierul uman

face acest lucru – probabil că suntem foarte încrezători în răspunsul nostru atunci când ni se cere numele mic al mamei, dar nu suntem la fel de încrezători în privința numelui unei persoane întâlnite ocazional în urmă cu un an.

Prin urmare, în loc să ofere o singură abordare elegantă pentru înțelegerea problemei de limbă inerentă în *Jeopardy!*, oamenii de știință de la IBM au combinat toate modulele de top în înțelegerea limbii pe care au putut să le găsească. Unele folosesc Modele Markov Ascunse Ierarhice; altele folosesc variante matematice ale Modelelor Markov Ascunse Ierarhice; iar altele utilizează abordări bazate pe reguli pentru codificarea unui set central de reguli sigure. UIMA evaluează performanțele fiecărui sistem în condiții propriu-zise de funcționare și le combină în mod optim. Există o doză de înțelegere greșită în discuțiile publice privitoare la Watson, sub aspectul că oamenii de știință de la IBM care l-au creat se concentrează adeseori asupra UIMA, managerul-expert produs de ei. Acest fapt conduce la comentarii de felul celui că Watson nu înțelege de fapt limba naturală deoarece este dificil să identifici în ce anume rezidă această înțelegere. Deși scheletul UIMA învață și din propria experiență, „înțelegerea" limbii de către Watson nu poate să se regăsească doar în UIMA, ci este distribuită în toate numeroasele sale componente, inclusiv în modulele autoorganizatoare de limbă care folosesc metode similare Modelelor Markov Ascunse Ierarhice.

O parte separată a tehnologiei lui Watson face apel la estimarea încrederii, care ține de UIMA, pentru a determina felul cum să plaseze pariurile la *Jeopardy!* În vreme ce sistemul Watson este special optimizat pentru a juca acest joc, tehnologia lui de bază pentru căutarea în limbă și cunoaștere poate fi adaptată cu ușurință în scopul îndeplinirii unor sarcini mai largi. Se poate crede că un areal de cunoștințe mai puțin obișnuit, cum este domeniul medical, ar fi mult mai ușor de stăpânit decât cunoștințele „obișnuite" necesare pentru a juca *Jeopardy!* De fapt, lucrurile stau exact invers: cunoștințele profesionale tind să fie mai bine organizate și structurate, deci mai puțin ambigue decât corespondentele lor comune, astfel încât folosirea acestor tehnici este deosebit de pretabilă la înțelegerea limbii naturale. După cum am menționat, IBM lucrează actualmente cu Nuance pentru adaptarea tehnologiei lui Watson la medicină.

Conversația care are loc atunci când Watson joacă *Jeopardy!* este scurtă: se pune o întrebare, iar Watson dă răspunsul. (Din nou, din punct de vedere tehnic, vine cu o întrebare care să se potrivească răspunsului propus în concurs.) Nu se angajează într-o conversație care ar impune urmărirea

tuturor afirmațiilor anterioare ale tuturor participanților. (Siri face acest lucru, dar într-o măsură redusă: dacă îi ceri să-i transmită soției tale un mesaj, îți va cere să o identifici și va reține cine este aceasta pentru solicitările ulterioare.) Urmărirea întregii informații din cadrul unei conversații – sarcină care ar fi în mod evident impusă pentru a trece Testul Turing – e o cerință suplimentară semnificativă, dar nu cu mult mai dificilă decât ceea ce face deja Watson. La urma urmelor, Watson a citit sute de milioane de pagini de material, iar acesta a cuprins, în mod evident, și numeroase relatări, astfel încât computerul este capabil să urmărească și evenimente cu succesiuni complexe. Prin urmare, ar trebui să fie capabil să susțină propriile conversații și să le ia în considerare în replicile ulterioare.

O altă limitare a jocului *Jeopardy!* rezidă în faptul că răspunsurile sunt în general scurte: de exemplu, nu se pun întrebări care să-i oblige pe concurenți să numească cele cinci teme principale ale romanului *Poveste despre două orașe*. În măsura în care poate găsi documente care să discute temele din acest roman, o versiune a lui Watson modificată corespunzător ar trebui să fie în stare să răspundă și la această întrebare. Faptul de a avansa asemenea teme în mod independent, numai în urma lecturării cărții și nu prin copierea gândurilor altor gânditori (chiar și fără cuvinte), reprezintă o altă chestiune. Această abordare este o altă sarcină de nivel superior pe care Watson se dovedește capabil s-o îndeplinească astăzi – e ceea ce aș numi sarcină de nivelul Testului Turing. (Acestea fiind spuse, voi scoate în evidență faptul că majoritatea oamenilor nu vin nici ei cu gânduri proprii și originale, ci copiază ideile colegilor și opiniile conducătorilor.) În orice caz, suntem în anul 2012, nu 2029, așa că, deocamdată, nici nu mă aștept la o inteligență de nivelul Testului Turing. Dar, pe de altă parte, aș scoate în evidență faptul că evaluarea răspunsurilor la solicitări cum ar fi găsirea ideilor-cheie dintr-un roman nu este, ca atare, o sarcină simplă. Dacă ești întrebat cine a semnat *Declarația de Independență*, poți să evaluezi dacă răspunsul tău este adevărat sau nu. Validitatea răspunsurilor la întrebări de nivel superior, cum ar fi descrierea temelor unei lucrări de creație, este mult mai greu de stabilit.

Demn de evidențiat este și faptul că, deși aptitudinile lingvistice ale sale sunt, pe undeva, inferioare celor ale unei persoane umane cu o educație solidă, Watson s-a arătat capabil să-i înfrângă pe cei mai mari doi jucători de *Jeopardy!* din lume. Și a reușit să realizeze asta deoarece poate combina abilitatea lingvistică și înțelegerea cunoștințelor cu utilizarea perfectă și de o remarcabilă acuratețe a memoriei pe care o posedă mașinile. Acesta este

motivul pentru care le-am încredințat mașinilor deja, într-o mare măsură, memoria noastră personală, socială și istorică.

Cu toate că nu sunt pregătit să-mi devansez predicția privitoare la un computer care să treacă Testul Turing înaintea anului 2029, progresele realizate în cazul unor sisteme ca Watson ar trebui să dea oricui o mare doză de încredere că apariția inteligenței artificiale de nivelul Testului Turing este aproape. Dacă ar trebui să se creeze o versiune a lui Watson optimizată pentru Testul Turing, probabil că s-ar ajunge foarte aproape de atingerea scopului.

Filosoful american John Searle (născut în anul 1932) a susținut recent că Watson nu este capabil să gândească. Citându-și propriul experiment mintal, denumit „Camera chinezească" (despre care voi discuta mai pe larg în Capitolul 11), el afirmă că Watson nu face decât să manipuleze simboluri și că nu înțelege semnificația acestor simboluri. Practic, Searle nu îl descrie pe Watson cu acuratețe, deoarece înțelegerea limbajului de către acesta din urmă se bazează pe procese statistice ierarhice – nu pe manipularea simbolurilor. Singura modalitate de a aprecia drept exactă caracterizarea lui Searle ar fi aceea de a considera fiecare pas al proceselor de autoorganizare ale lui Watson ca fiind „manipulare a simbolurilor". Dar, într-un asemenea caz, nici creierul omului nu ar putea fi considerat capabil de judecată.

Este amuzant și de râs faptul că observatorii îl critică pe Watson că nu face altceva decât o analiză statistică a limbii, în loc să proceseze „adevărata" înțelegere a limbii, care îi este proprie omului. Dar creierul uman efectuează exact o analiză statistică ierarhică atunci când rezolvă ipoteze multiple pe baza deducerii statistice (și realmente la fiecare nivel al ierarhiei neocortexului). Atât Watson, cât și creierul omului învață și răspund pe baza unei abordări similare înțelegerii ierarhice. Din multe puncte de vedere, cunoștințele lui Watson sunt mult mai cuprinzătoare decât ale omului; niciun om nu poate să pretindă că stăpânește toată Wikipedia, care reprezintă doar o parte din baza de cunoștințe a lui Watson. Invers, astăzi omul poate stăpâni mai multe niveluri conceptuale decât Watson, dar în mod sigur deosebirea nu este definitivă.

Un sistem important care demonstrează forța computației aplicată cunoștințelor organizate este Wolfram Alpha, un motor de răspunsuri (diferit de un motor de căutare) pus la punct de matematicianul și omul de știință britanic dr. Wolfram (născut în anul 1959), împreună cu colegii săi de la Wolfram Research. De exemplu, dacă îl întrebi pe Wolfram Alpha (la

WolframAlpha.com) „Câte numere prime sunt într-un milion?", acesta îți va răspunde „78.498". N-a căutat numărul acesta, ci l-a computat şi, după răspuns, îți va da şi ecuațiile pe care le-a folosit. Dacă ai încerca să ajungi la răspuns folosind un motor de căutare convențional, acesta te-ar orienta către link-uri în care ai putea găsi algoritmii necesari. Pe urmă, ar trebui să introduci aceşti algoritmi într-un sistem cum ar fi *Mathematica*, pus la punct tot de dr. Wolfram, însă această abordare ți-ar solicita în mod evident mult mai mult timp (şi capacitate de înțelegere) decât cel(e) necesar(e) ca să-l întrebi pur şi simplu pe Alpha.

Într-adevăr, Alpha constă în 15 milioane de linii de cod *Mathematica*. El poate computa răspunsul din aproximativ 10 mii de miliarde de date care au fost atent alese de echipa de la Wolfram Research. Poți să-i adresezi lui Alpha o gamă largă de întrebări efective, cum ar fi: „Ce țară are cel mai mare PIB pe locuitor?" (Răspuns: „Monaco, 212.000 de dolari americani pe cap de locuitor") sau „Ce vârstă are Stephen Wolfram?" (Răspuns: „52 de ani, 9 luni şi 2 zile" în momentul în care scriu aceste rânduri). După cum am menționat, Alpha este folosit ca parte a lui Siri de la Apple; dacă îi adresezi lui Siri o întrebare referitoare la fapte reale, aceasta îi este transmisă mai departe lui Alpha, pentru a răspunde la ea. De asemenea, Alpha se ocupă de căutările solicitate pe motorul de căutare Bing al lui Microsoft.

Într-o postare recentă, dr. Wolfram a anunțat că Alpha dă actualmente răspunsuri corecte în 90 la sută dintre cazuri.[17] De asemenea, a comunicat o reducere exponențială a ratei eşecurilor, cu un timp de înjumătățire de circa optsprezece luni. Este un sistem impresionant şi foloseşte metode create manual şi date verificate manual. E o explicație a motivului pentru care au fost create de fapt computerele. În timp ce noi descoperim şi compilăm metode ştiințifice şi matematice, computerele se dovedesc a fi cu mult mai bune la implementarea lor decât inteligența umană neasistată. Cele mai multe metode ştiințifice au fost codificate în Alpha laolaltă cu date actualizate permanent pe subiecte ce merg de la economie la fizică. În cursul unei conversații particulare pe care am purtat-o cu el, dr. Wolfram a estimat că metodele autoorganizatoare de felul celor folosite în cazul lui Watson ating o precizie de aproximativ 80 la sută atunci când funcționează bine. Alpha, a menționat dr. Wolfram, ajunge la o precizie de aproximativ 90 la sută. Desigur, în cazul ambelor date procentuale de mai înainte, intervine o autoselecție, ceea ce înseamnă că utilizatorii (bunăoară, eu) au învățat genul de întrebări la care este bun Alpha. Un proces similar se aplică şi metodelor

autoorganizatoare. Prin urmare, optzeci la sută pare să fie o estimare rezonabilă pentru precizia lui Watson la întrebările de la *Jeopardy!*, iar ea a fost suficient de mare pentru ca Watson să-i înfrângă pe oamenii cei mai buni.

Opinia mea este că metodele autoorganizatoare de felul celor despre care am vorbit în cadrul Teoriei minții bazate pe recunoașterea modelelor sunt necesare pentru înțelegerea ierarhiilor complexe și adeseori ambigue cu care ne confruntăm în cazul fenomenelor din lumea reală – inclusiv limba omului. O combinație ideală pentru un sistem inteligent robust ar presupune reunirea inteligenței ierarhice susținute de Teoria minții bazate pe recunoașterea modelelor (care afirmă că acesta este modul de funcționare a creierului omului) cu codificarea precisă a cunoștințelor și datelor științifice. În esență, asta apropie omul de computer. Vom dezvolta ambii poli ai inteligenței în anii care urmează. În ceea ce privește inteligența noastră biologică, deși neocortexul prezintă o plasticitate semnificativă, arhitectura lui de bază este limitată de constrângerile fizice. Apariția unui neocortex suplimentar în fruntea omului a fost o inovație importantă a evoluției, însă acum nu mai putem să ne mărim lobii frontali de o mie de ori, ba poate nici măcar cu 10 la sută. Cu alte cuvinte, nu putem realiza acest lucru în mod biologic, însă îl vom realiza prin mijloace tehnologice.

O strategie pentru crearea unei minți

> În creierul nostru există miliarde de neuroni, dar ce sunt neuronii? Niște simple celule. Creierul nu are niciun fel de cunoștințe înainte de a se realiza conexiunile dintre neuroni. Tot ce știm, tot ce suntem vine din felul în care sunt conectați neuronii.
>
> **— Tim Berners-Lee**

Să folosim observațiile pe care le-am discutat anterior pentru a începe să construim un creier. Vom porni de la realizarea unui recognizator de modele care să satisfacă toate cerințele impuse. Pe urmă, vom face còpii ale recognizatorului, în măsura în care pot să suporte memoria și capacitățile noastre de computație. Orice recognizator computează probabilitatea ca modelul lui să fie recunoscut. Procedând astfel, ia în considerare magnitudinea observată a fiecărui input (în cadrul unui continuum adecvat) și o compară cu parametrii de mărime și de variabilitate a mărimii asociați

fiecărui input. Recognizatorul își excită axonul simulat, dacă probabilitatea computată depășește un anumit prag. Acest prag și parametrii care controlează computația probabilității modelului se numără printre elementele pe care le vom optimiza cu ajutorul unui algoritm genetic. Cum nu există cerința ca fiecare input să fie activ pentru a fi recunoscut un anumit model, avem o recunoaștere autoasociativă (ceea ce înseamnă recunoașterea unui model pe baza prezentării unei părți a acelui model). De asemenea, permitem emiterea de semnale inhibitoare (semnale ce indică faptul că e mai puțin probabil să fie vorba de modelul respectiv). Recunoașterea modelului determină generarea unui semnal activ pe axonul simulat al acestui recognizator de modele. Axonul este, la rândul lui, conectat cu unul sau mai mulți recognizatori de modele de la nivelul conceptual superior. Toți recognizatorii de modele conectați la următorul nivel conceptual superior acceptă acest model ca fiind unul dintre input-urile lor. Fiecare recognizator trimite semnale și la recognizatorii de modele de nivel conceptual inferior, de câte ori recunoaște un model, transmițându-le că este „de așteptat" restul modelului. Fiecare recognizator de modele are unul sau mai multe asemenea canale de input „de așteptat". Când este primit, astfel, semnalul „de așteptat", pragul de recunoaștere pentru acest recognizator de modele e diminuat (e făcut mai accesibil).

Recognizatorii de modele sunt răspunzători pentru „legarea" la alți recognizatori de modele, superiori sau inferiori, din ierarhia conceptuală. A se reține că toate „legăturile" în cazul unei implementări software operează prin intermediul link-urilor virtuale (care, asemenea link-urilor web, sunt în esență indicatoare de memorie) și nu prin legături propriu-zise. Acest sistem este, practic, mai flexibil decât cel din creierul biologic. În creierul uman, modelele noi trebuie să fie alocate unui recognizator de modele real, fizic, iar conexiunile noi trebuie realizate prin intermediul legăturii obișnuite axon-dendrită. De obicei, asta înseamnă că se ia o conexiune fizică existentă, adică elementul de care e nevoie, iar apoi se măresc extensiile necesare ale axonului și dendritelor pentru realizarea conectării depline.

O altă tehnică folosită de creierul biologic al mamiferelor este să pornească de la un număr mare de conexiuni posibile și apoi să taie selectiv conexiunile neuronale care nu sunt folosite. Dacă un neocortex biologic realocă recognizatori corticali de modele care au învățat deja modele mai vechi pentru a asimila un material nou, conexiunile trebuie reconfigurate fizic. Din nou, aceste sarcini sunt mult mai simple în cazul unei implementări

software. Pur şi simplu alocăm noi poziţii în memorie unui nou recognizator de modele şi folosim link-urile de memorie pentru conexiuni. Dacă neocortexul digital vrea să transfere resurse ale memoriei corticale de la un set de modele la altul, pur şi simplu returnează memoriei vechii recognizatori de modele şi apoi face noua alocare. Acest gen de „adunare a gunoiului" şi realocare a memoriei este o caracteristică-standard a arhitecturii multor sisteme software. În creierul nostru digital, am salvat şi amintirile vechi înainte să le descărcăm din neocortexul activ – măsură de prevedere pe care nu o putem lua în cazul creierului biologic.

Există o multitudine de tehnici matematice care pot fi folosite pentru implementarea acestei abordări privind recunoaşterea ierarhică autoorganizată a modelelor. Metoda pe care aş folosi-o este aceea a Modelelor Markov Ascunse – şi asta din mai multe motive. Din perspectiva mea, am adunat câteva decenii de familiarizare cu această metodă, folosind-o la sistemele de pionierat de recunoaştere a vorbirii şi a limbii naturale, începând cu anii 1980. Din perspectiva domeniului în ansamblul său, există pentru Modelele Markov Ascunse o experienţă mai cuprinzătoare decât pentru orice altă abordare a sarcinii de recunoaştere a modelelor. Modelele Markov Ascunse sunt folosite considerabil şi în înţelegerea limbii naturale. Numeroase sisteme de înţelegere a limbii naturale folosesc tehnici cel puţin matematic similare Modelelor Markov Ascunse Ierarhice.

A se observa că nu toate sistemele bazate pe Modele Markov Ascunse sunt complet ierarhice. Unele permit existenţa a doar câteva niveluri ale ierarhiei – de exemplu, trecerea de la stările acustice la foneme şi apoi la cuvinte. Pentru construirea unui creier, vom dori să-i permitem sistemului nostru să creeze oricâte nivele ierarhice îi sunt necesare. De asemenea, majoritatea sistemelor bazate pe Modele Markov Ascunse nu sunt complet autoorganizate. Unele au conexiuni fixe, deşi aceste sisteme îşi taie efectiv multe dintre conexiunile de început, permiţându-le să ajungă la greutăţi de conexiune zero. Sistemele noastre din anii 1980 şi 1990 îşi tăiau automat conexiunile cu greutăţi aflate sub un anumit nivel şi, de asemenea, permiteau stabilirea unor noi conexiuni pentru a crea modele ale datelor de pregătire în condiţii mai bune şi pentru a învăţa din mers. O cerinţă-cheie, cred eu, este să i se permită sistemului să-şi creeze flexibil propriile topologii, pe baza modelelor la care este expus în cursul învăţării. Putem folosi tehnica matematică a programării liniare în scopul alocării optime a conexiunilor pentru noii recognizatori de modele.

Creierul nostru digital va acorda, de asemenea, o redundanţă substanţială fiecărui model, în special celor care apar frecvent. Aceasta permite o recunoaştere robustă a modelelor obişnuite şi este, de asemenea, una dintre metodele-cheie pentru asigurarea unei recunoaşteri invariabile a diferitelor forme ale unui model. Pe de altă parte, vom avea nevoie de reguli cu privire la nivelul de redundanţă care poate fi admis, deoarece nu avem interesul să folosim un volum excesiv de memorie pentru fiecare model obişnuit, de nivel redus.

Regulile referitoare la redundanţă, la pragurile de recunoaştere şi la efectul asupra pragului unei indicaţii de genul „Acest model este de aşteptat" sunt câteva dintre exemplele de parametri-cheie de ansamblu care afectează performanţa sistemului cu autoorganizare. Iniţial, aş stabili aceşti parametri pe baza intuiţiei mele, dar i-aş optimiza apoi folosind un algoritm genetic.

Un considerent foarte important îl reprezintă educaţia creierului, indiferent dacă e vorba despre unul biologic sau de unul software. După cum am discutat anterior, un sistem ierarhic de recunoaştere a modelelor (digital sau biologic) va învăţa despre două – preferabil unul singur – niveluri ierarhice odată. Pentru a amorsa sistemul, aş porni de la reţele ierarhice pregătite anterior, care şi-au învăţat deja lecţiile de recunoaştere a vorbirii umane, a caracterelor tipărite şi a structurilor de limbă naturală. Un asemenea sistem ar fi capabil să citească documente în limbaj natural, dar nu ar putea să stăpânească la un moment dat decât un nivel conceptual aproximativ. Nivelurile asimilate anterior ar asigura o bază relativ stabilă pentru învăţarea următorului nivel. Sistemul poate citi acelaşi document în repetate rânduri, atingând noi niveluri conceptuale cu fiecare citire ulterioară, similar modului în care oamenii recitesc şi ajung la o înţelegere mai profundă a textelor. Pe web sunt disponibile miliarde de pagini de materiale. Wikipedia însăşi conţine în jur de patru milioane de articole în versiune engleză.

Aş oferi, de asemenea, un modul de gândire critică având proprietatea de a realiza o scanare continuă de fundal a tuturor modelelor existente, revizuindu-le compatibilitatea cu alte modele (idei) în acest neocortex software. Nu avem o asemenea facilitate în creierele noastre biologice, motiv pentru care oamenii pot să nutrească în mod egal gânduri complet inconsecvente. La identificarea unei idei inconsecvente, modulul digital începe să caute o rezolvare, incluzând propriile structuri corticale şi întreaga literatură care îi stă la dispoziţie. O decizie poate să însemne, de pildă, determinarea faptului că una dintre ideile inconsecvente este pur şi simplu incorectă (dacă faptul e

contraindicat de o preponderență a datelor aflate în conflict). Mai constructiv, ar găsi la un nivel conceptual superior o idee capabilă să rezolve aparenta contradicție, asigurând o perspectivă ce explică fiecare idee. Sistemul ar adăuga această rezoluție ca nou model şi ca link la ideile care au declanşat căutarea soluției. Acest modul critic de gândire ar sluji ca funcție continuă de fundal. Ar fi foarte benefic dacă creierul uman ar face acelaşi lucru.

De asemenea, ar asigura un modul de identificare a întrebărilor deschise din fiecare disciplină. Ca altă funcție continuă de fundal, ar căuta soluții pentru respectivele întrebări în alte arii ale cunoaşterii. După cum am subliniat, la nivelul neocortexului, cunoştințele constau în module profund înrădăcinate ale modelelor şi, prin urmare, sunt complet metaforice. Un model nu poate fi folosit decât pentru a asigura o soluție, ori ca sursă de intuiție pentru un domeniu aparent fără legătură.

Ca exemplu, aminteşte-ți de metafora pe care am folosit-o în Capitolul 4, cea care lega mişcările aleatorii ale moleculelor într-un gaz de mişcările aleatorii ale schimbărilor evolutive. Moleculele dintr-un gaz se mişcă aleatoriu, fără niciun simț aparent al direcției. În ciuda acestui fapt, dacă i se va acorda suficient timp, practic fiecare moleculă dintr-un gaz aflat într-un pahar de laborator va părăsi recipientul. Am observat că acest fapt asigură o perspectivă legată de o întrebare importantă referitoare la evoluția inteligenței. Asemenea moleculelor de gaz, schimbările evolutive se îndreaptă în toate părțile, aparent fără direcție. Cu toate acestea, sesizăm o mişcare către o mai mare complexitate şi o mai mare inteligență, către realizarea supremă a evoluției unui neocortex capabil de gândire ierarhică. Aşadar, suntem capabili să ne facem o idee despre modul în care un proces aparent lipsit de scop şi de direcție poate avea un rezultat aparent plin de logică într-un domeniu (evoluția biologică), îndreptându-ne atenția către un alt domeniu (termodinamica).

Am menționat mai devreme că ideea lui Charles Lyell privitoare la faptul că modificările minore provocate de apele curgătoare în corpul formațiunilor stâncoase pot duce la apariția unor văi mari l-a inspirat peste timp pe Charles Darwin să facă o observație similară în legătură cu modificările minime continue ale caracteristicilor organismelor în cadrul unei specii. Această căutare metaforică ar fi un alt proces continuu de fundal.

Ar trebui să asigurăm un mod de parcurgere a mai multor liste în mod simultan, pentru garantarea echivalentului gândului structurat. O listă poate însemna afirmarea constrângerilor pe care trebuie să le satisfacă soluția la

o problemă. Fiecare pas poate genera o căutare recursivă în ierarhia existentă de idei sau o căutare prin literatura disponibilă. Creierul uman pare a fi capabil să facă față doar unui număr de patru liste simultan (fără ajutorul unor instrumente precum computerul), dar nu există niciun motiv pentru care un neocortex artificial să aibă o asemenea limitare.

De asemenea, vom vrea să ne mărim creierul artificial cu genul de inteligență în care au excelat întotdeauna computerele, asta însemnând capacitatea de a stăpâni cu precizie vaste baze de date și de a implementa algoritmi cunoscuți în mod rapid și eficient. Wolfram Alpha combină într-un fel unic numeroase metode științifice cunoscute și le aplică unor date adunate atent. Acest tip de sistem va continua de asemenea să se îmbunătățească, în virtutea observației dr. Wolfram referitoare la un declin exponențial al ratei erorilor.

În cele din urmă, creierul nostru are nevoie de scop. Scopul este exprimat ca o serie de scopuri. În cazul creierului biologic, scopurile ne sunt stabilite de centrii de plăcere și de frică pe care i-am moștenit de la creierul vechi. Aceste porniri primitive au fost generate inițial de evoluția biologică pentru asigurarea supraviețuirii speciilor, dar neocortexul ne-a permis să le sublimăm. Scopul lui Watson a fost să răspundă la întrebările de la *Jeopardy!* Un alt scop formulat simplu poate să fie trecerea Testului Turing. Pentru asta, creierul digital are nevoie de nararea umană a unei poveşti fictive proprii, astfel încât să poată pretinde că este uman din punct de vedere biologic. De asemenea, el trebuie să se „prostească" în mod considerabil, căci orice sistem care ar manifesta cunoştințele lui Watson, să spunem, ar fi demascat repede ca fiind unul nonbiologic.

Fapt foarte interesant, am putea să-i dăm creierului nostru un scop mai ambițios, ca de pildă acela de a contribui la o lume mai bună. Desigur, un scop de acest fel ridică o mulțime de întrebări: Mai bună pentru cine? Mai bună în ce fel? Pentru omul biologic? Pentru ființele conştiente? În acest caz, cine sau ce înseamnă conştient?

Pe măsură ce creierul nonbiologic devine tot atât de capabil precum cel biologic să efectueze schimbări în lume – pe bună dreptate, în cele din urmă mult mai capabil decât creierul biologic neperfecționat –, va trebui să luăm în considerare educarea lui morală. Un bun punct de pornire l-ar putea reprezenta o veche idee din tradițiile noastre religioase: Regula de Aur.

Capitolul 8

MINTEA CA UN COMPUTER

Având forma aproximativă a unei felii de pâine țărănească din Franța, creierul nostru este un aglomerat laborator de chimie, care debordează nonstop de conversații neuronale. Imaginează-ți creierul: acel ghem lucios al existenței, acel parlament de celule gri-şoarece, acea fabrică de vise, acel mic tiran plasat într-o bilă osoasă, acel talmeş-balmeş de neuroni care comandă toate actele, acel mic pretutindeni, acel capricios teren de distracție, acea garderobă încrețită de euri înghesuite în craniu, la fel ca prea multe haine într-o geantă pentru sală.

— Diane Ackerman

Creierul există pentru că distribuția de resurse necesare supraviețuirii şi pericolele care amenință supraviețuirea variază în spațiu şi timp.

— John M. Allman

Geografia modernă a creierului are o calitate delicios antichizată – foarte asemănătoare unei hărți cu lumea cunoscută încercuită de *terra incognita* în care bântuie monştri.

— David Bainbridge

În matematică, nu trebuie să înțelegi lucrurile. Pur şi simplu te obişnuieşti cu ele.

— John von Neumann

Încă de la apariția computerului, la mijlocul secolului douăzeci, a avut loc o dezbatere continuă nu numai pe tema extinderii finale a abilităților acestuia, ci şi dacă nu cumva însuşi creierul uman poate fi considerat o formă de computer. În ceea ce priveşte a doua temă, consensul a virat de la considerarea acestor două genuri de entități care procesează informații ca fiind esențialmente identice la opinia că sunt fundamental diferite. Prin urmare, este creierul un computer?

Când au devenit pentru prima oară un subiect popular, în anii 1940, computerele au fost considerate imediat mașini gânditoare. Computerul ENIAC, anunțat în anul 1946, a fost descris în presă ca fiind un „creier gigantic". Pe măsură ce computerele au devenit disponibile comercial în cursul următorului deceniu, publicitatea s-a referit în mod curent la ele ca la niște creiere capabile de roluri cărora creierul obișnuit, biologic, nu le poate face față.

O reclamă din 1957 ce prezintă punctul de vedere popular, care considera computerul drept un creier gigantic

Programele de computer le-au permis mașinilor să se prezinte la înălțimea acestei aprecieri. „Rezolvitorul general de probleme", creat în anul 1959 de Herbert A. Simon, J.C. Shaw și Allen Newell la Universitatea Carnegie, a reușit să facă demonstrația unei teoreme pe care matematicienii Bertrand Russell (1872-1970) și Alfred North Whitehead (1861-1947) nu reușiseră să o demonstreze în celebra lor lucrare din anul 1913 *Principia Mathematica*. A devenit evident în cursul deceniilor care au urmat că aceste computere puteau să depășească repede și semnificativ capacitățile umane

neasistate în cazul unor exerciții intelectuale cum ar fi rezolvarea problemelor matematice, diagnosticarea bolilor şi jocul de şah, însă aveau dificultăți în dirijarea unui robot să lege şireturile de la pantof sau în înțelegerea limbii obişnuite pe care putea să o priceapă un copil de cinci ani. Computerele abia încep acum să stăpânească acest gen de abilități. În mod ironic, evoluția inteligenței computerului a avut o direcție opusă maturizării omului.

Problema dacă creierul uman şi computerul sunt, la un anumit nivel, echivalente rămâne controversată şi azi. În *Introducere* am menționat că există milioane de link-uri de citate pe tema complexității creierului uman. În mod similar, o căutare Google pentru „Citate: Creierul nu este un computer" dă de asemenea milioane de link-uri. Din punctul meu de vedere, afirmațiile pe această temă sunt echivalente cu a spune „Sucul concentrat de măr nu e măr". Sub aspect tehnic, această afirmație este adevărată, dar poți să faci suc concentrat de măr dintr-un măr. Poate şi mai aproape de subiect, asta este ca şi cum ai spune „Computerele nu sunt procesoare de cuvinte". Este adevărat că atât computerul, cât şi procesorul de cuvinte există la nivele conceptuale diferite, dar un computer poate deveni procesor de cuvinte dacă rulează un software de procesare a acestora, nu şi invers. În mod similar, computerul poate deveni creier dacă rulează software de creier. Acesta este lucrul pe care vor să-l facă cercetătorii, inclusiv eu.

Şi atunci, întrebarea este dacă putem sau nu să găsim un algoritm care să transforme computerul într-o entitate echivalentă cu un creier uman. La urma urmelor, computerul poate rula orice algoritm pe care îl putem defini, datorită universalității lui înnăscute (supuse numai capacității lui). Creierul uman, pe de altă parte, funcționează cu un set specific de algoritmi. Metodele lui sunt inteligente prin faptul că permit o semnificativă plasticitate şi restructurarea propriilor conexiuni pe baza experienței sale, însă aceste funcții pot fi imitate de un software.

Universalitatea computației (ideea că un computer de uz general poate să implementeze orice fel de algoritm) şi forța acestei idei au apărut în acelaşi timp cu primele maşini propriu-zise. Există patru concepte-cheie care stau la baza universalității, fezabilității şi aplicabilității computației la gândirea umană. Ele merită să fie trecute în revistă aici, deoarece creierul însuşi le foloseşte. Primul îl reprezintă abilitatea de comunicare, reținere şi computare a informației într-un mod sigur. În jurul anului 1940, dacă foloseai cuvântul „computer", oamenii presupuneau că vorbeai despre un computer analogic, la care numerele erau reprezentate de diferite tensiuni electrice, iar

componentele specializate puteau să realizeze operații matematice, cum ar fi adunarea și înmulțirea. O mare limitare a computerelor analogice a constat în faptul că erau marcate de probleme de acuratețe. Numerele puteau fi reprezentate cu o precizie de numai o sutime, iar când tensiunile electrice care le reprezentau erau procesate prin creșterea numărului de operatori matematici, erorile se acumulau. Dacă voiai să realizezi ceva mai mult de o mână de operații, rezultatele deveneau atât de imprecise, încât își pierdeau sensul.

Oricine își mai amintește de vremurile în care muzica era înregistrată pe mașini cu bandă analogice își va aduce aminte și de acest efect. La prima copie apărea o degradare sesizabilă, conținând ceva mai mult zgomot decât originalul. (Îți amintești, dragă cititorule, că „zgomotul" reprezintă impreciziile aleatorii.) O copie a copiei conținea și mai mult zgomot, iar până la a zecea generație, copia însemna aproape complet numai zgomot. S-a presupus că aceeași problemă avea să marcheze și apariția lumii computerelor digitale. Putem înțelege asemenea preocupări dacă luăm în calcul comunicarea informației digitale printr-un canal. Niciun canal nu este perfect și fiecare prezintă o anumită rată de eroare. Să presupunem că avem un canal cu probabilitatea 0,9 de transmitere corectă a fiecărui bit. Dacă aș expedia un mesaj cu lungimea de un bit, probabilitatea transmiterii lui exacte pe acest canal ar fi de 0,9. Presupunem că transmitem doi biți? Acum, acuratețea este de $0,9^2 = 0,81$. Dar dacă aș trimite un byte (opt biți)? Am o șansă nu tocmai dreaptă (0,43, ca să fiu exact) de transmitere corectă a acestui byte. Probabilitatea expedierii precise a unui volum de informație de cinci byte este de aproximativ 1 la sută.

O soluție evidentă pentru ocolirea problemei este ca acest canal să devină mai precis. Să presupunem că respectivul canal face doar o eroare la un milion de biți. Dacă aș transmite un fișier constând în jumătate de milion de bytes (cam de proporțiile unui program modest sau ale unei baze de date), probabilitatea expedierii corecte ar fi mai mică de 2 la sută, în ciuda acurateței foarte mari a canalului. Dat fiind faptul că, în cazul unui singur bit, eroarea poate invalida complet un program de computer și alte tipuri de date digitale, situația nu este satisfăcătoare. Indiferent de acuratețea canalului, deoarece probabilitatea apariției unei erori într-o transmisie crește rapid odată cu volumul mesajului, s-ar putea să pară că avem de-a face cu o barieră refractară.

Computerele analogice au abordat această problemă printr-o degradare grațioasă (ceea ce însemna că utilizatorii prezentau doar probleme în care

se puteau tolera mici erori); însă dacă utilizatorii computerelor analogice se limitau la un set constrâns de calcule, computerele se dovedeau pe undeva utile. Computerele digitale, pe de altă parte, necesită o comunicare continuă, nu numai de la un computer la altul, ci şi la nivelul aceluiaşi computer. Aici există comunicare de la şi către memoriile sale, către şi de la unitatea de procesare. În unitatea centrală de procesare, există comunicare de la un registru la altul şi către şi dinspre unitatea aritmetică şi aşa mai departe. Chiar şi în cazul unităţii aritmetice există comunicare între un registru de biţi şi altul. Comunicarea este atotprezentă la toate nivelurile. Dacă ţinem seama de faptul că rata erorilor creşte rapid odată cu înmulţirea comunicaţiilor şi că o eroare de un singur bit poate distruge integritatea unui proces, computaţia digitală era condamnată – sau aşa părea la vremea respectivă.

Fapt remarcabil, aceasta a fost perspectiva obişnuită până când a apărut matematicianul american Claude Shannon (1916–2001) şi a demonstrat cum putem crea în mod arbitrar o comunicare precisă folosind în acest scop chiar şi cele mai nesigure canale. Ce a afirmat Shannon în lucrarea lui fundamentală *A Mathematical Theory of Communication* – *O teorie matematică a comunicării*, publicată în *Bell System Technical Journal* în lunile iulie şi octombrie 1948 şi în special în teorema lui privind codificarea pentru canale zgomotoase a fost că, dacă ai la dispoziţie un canal cu o rată de eroare oarecare (cu excepţia celor de exact 50 la sută pe bit, fapt ce ar însemna că ele ar transmite doar zgomot pur), poţi expedia un mesaj pentru care rata de eroare va fi cât de redusă vrei tu. Cu alte cuvinte, rata de eroare a transmisiunii poate fi de un bit din *n* biţi, unde *n* poate fi oricât de mare vrei tu să fie. Aşa, de exemplu, la extrem, dacă ai un canal care transmite în mod corect biţi de informaţie în numai 51 la sută din timp (respectiv, transmite bitul corect doar cu ceva mai multă frecvenţă decât bitul greşit), poţi totuşi să expediezi mesaje în care un singur bit dintr-un milion să fie incorect sau chiar unul singur dintr-o mie de miliarde sau dintr-un miliard de miliarde.

Cum e posibil aşa ceva? Răspunsul este: prin redundanţă. Acest lucru poate părea evident acum, dar nu a fost aşa şi la momentul respectiv. Ca exemplu simplu, dacă transmit fiecare bit de trei ori şi mă bazez pe votul majorităţii, sporesc substanţial fiabilitatea. Dacă nici aşa nu e suficient de bine, pur şi simplu măresc redundanţa până obţin fiabilitatea pe care mi-o doresc. Simpla repetare a informaţiei reprezintă modul cel mai simplu de asigurare a unor rate de acurateţe ridicate pe canale de acurateţe redusă, dar ea nu este şi cea mai eficientă abordare. Lucrarea lui Shannon, care a

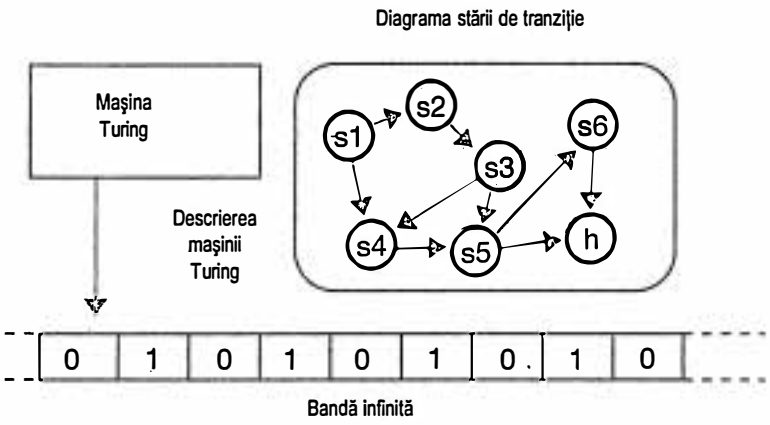
întemeiat domeniul teoriei informației, prezenta și metode optime sau coduri de detectare și corectare a erorilor care pot asigura orice grad de acuratețe pe orice canal nealeatoriu.

Cititorii mai vârstnici își amintesc de modemurile telefonice, care transmiteau informația pe linii analogice cu nivel ridicat de zgomot. Aceste linii aveau un fond sonor cu șuierături și pocnete evidente auditiv, precum și cu alte forme de distorsiuni, dar, în ciuda acestui fapt, reușeau să transmită datele digitale cu rate foarte înalte de acuratețe, mulțumită teoremei canalului zgomotos a lui Shannon. Aceeași problemă și aceeași soluție există și pentru memoria digitală. Te-ai întrebat vreodată, cititorule, cum reușesc CD-urile, DVD-urile și discurile cu programe să dea rezultate fiabile chiar și după ce discul a fost scăpat pe jos sau zgâriat? Din nou ne putem gândi la Shannon.

Computația constă în trei elemente: comunicația – care, după cum am menționat, este atotprezentă atât în interiorul, cât și între computere –, memoria și porțile logice (care îndeplinesc funcțiile aritmetice și logice). Acuratețea porților logice poate fi de asemenea mărită, prin utilizarea similară a codurilor de detectare și corectare a erorilor. Datorită teoremei și teoriei lui Shannon putem noi acum să manevrăm volume arbitrar de mari sau complexe de date digitale și algoritmi fără ca procesele să fie tulburate sau distruse de erori. Este important să evidențiez faptul că creierul folosește și el principiul lui Shannon, deși evoluția celui uman în mod clar îl antedatează pe Shannon! Majoritatea modelelor sau ideilor (și o idee este tot un model), după cum am văzut, sunt stocate în creier cu un volum substanțial de redundanță. Un motiv primar care explică redundanța creierului este nefiabilitatea inerentă a circuitelor neuronale.

A doua idee importantă pe care se bazează epoca informației este cea menționată anterior: universalitatea computației. În anul 1936, Alan Turing și-a descris „mașina Turing", care nu era propriu-zis o mașină, ci un alt experiment mintal. Computerul lui teoretic constă într-o bandă de memorie de lungime infinită cu un 1 sau 0 pe fiecare pătrat. Input-ul mașinii este prezent pe această bandă, pe care mașina o poate citi pătrat cu pătrat. Mașina mai conține și un tabel de reguli – în esență, un program stocat – care constă în stări numărate. Fiecare regulă specifică o acțiune dacă pătratul citit pe moment este un 0 și o acțiune diferită dacă pătratul curent este 1. Posibilele acțiuni includ scrierea unui 0 sau 1 pe bandă, deplasarea benzii cu un pătrat la dreapta sau la stânga și oprirea ei. Fiecare stare specifică, prin urmare, numărul stării următoare în care ar trebui să se afle mașina.

Input-ul maşinii Turing este prezentat pe bandă. Programul rulează şi, când maşina se opreşte, şi-a încheiat algoritmul, iar output-ul procesului este lăsat pe bandă. A se observa că, deşi banda are, teoretic, o lungime infinită, orice program propriu-zis care nu se încadrează într-o buclă infinită va folosi doar o porţiune finită din bandă, aşadar, şi dacă ne limităm la o bandă finită, maşina tot va rezolva un set util de probleme.



O diagramă schematică a unei maşini Turing, cu capul care citeşte şi scrie pe bandă şi cu un program intern care constă în tranziţii de stare

Dacă maşina Turing pare simplă, asta se datorează faptului că acesta a fost obiectivul inventatorului ei. Turing a vrut ca maşina lui să fie cât mai simplă cu putinţă (dar nu mai simplă, ca să-l parafrazez pe Einstein). Turing şi Alonzo Church (1903-1995), fostul său profesor, au continuat şi au pus la punct Teza Church-Turing, care afirmă că, dacă o problemă ce poate fi pusă maşinii Turing nu poate fi rezolvată de aceasta, nu poate fi rezolvată de nicio altă maşină, potrivit legii naturale. Cu toate că maşina Turing are doar o mână de comenzi şi procesează doar câte un bit o dată, este capabilă să computeze orice pot computa alte computere. Un alt mod de a spune acest lucru este că orice maşină „completă Turing" (cu alte cuvinte, care are capacităţi echivalente cu o maşină Turing) poate computa orice algoritm (orice procedură pe care o putem defini).

Interpretările „dure" ale Tezei Church-Turing propun o echivalență esențială între ceea ce poate să gândească sau să știe un om și ceea ce poate computa o mașină. Ideea de bază este că creierul uman e supus și el legii naturale și, prin urmare, capacitatea lui de a procesa informația nu o poate depăși pe cea a unei mașini (așadar nici pe a mașinii Turing).

Îl putem credita pe bună dreptate pe Turing cu stabilirea fundamentului teoretic al computației odată cu apariția lucrării lui din anul 1936, dar este important de notat faptul că savantul a fost profund influențat de o prelegere pe care matematicianul american de origine maghiară John von Neumann (1903-1957) a ținut-o la Cambridge în anul 1935 pe tema conceptului său de program stocat, concept consfințit de mașina Turing.[1] La rândul lui, von Neumann a fost influențat de lucrarea din anul 1936 a lui Turing, care a stabilit cu eleganță principiile computației și a devenit o lectură obligatorie pentru colegii săi de la sfârșitul anilor 1930 și începutul anilor 1940.[2]

În aceeași lucrare, Turing comunică o altă descoperire neașteptată: aceea a problemelor insolvabile. Acestea sunt probleme bine definite, cu răspunsuri unice, care pot fi demonstrate că există, dar pe care putem dovedi că nu pot fi niciodată computate de vreo mașină Turing – cu alte cuvinte de nicio mașină, reversul a ceea ce a reprezentat o dogmă a secolului al nouăsprezecelea. Potrivit acesteia, problemele care pot fi definite vor fi în cele din urmă și rezolvate. Turing a arătat că există tot atâtea probleme nerezolvabile câte probleme rezolvabile există. Matematicianul și filosoful american de origine austriacă Kurt Gödel a ajuns la o concluzie similară în a sa „Teoremă a incompletitudinii" din anul 1936. Rămânem astfel în situația perplexantă de a fi capabili să stabilim o problemă, de a dovedi că există un răspuns unic, dar și de a ști că, totuși, răspunsul nu poate fi găsit vreodată.

Turing arătase că, în esența ei, computația se bazează pe un mecanism foarte simplu. Deoarece mașina Turing (și orice computer, prin urmare) este capabilă să își întemeieze viitorul curs de acțiune pe rezultatele pe care le-a computat deja, ea e capabilă și să ia decizii și să realizeze modele de complexitate arbitrară a ierarhiilor de informații.

În anul 1939, Turing a proiectat un calculator electronic numit *Bombe*, care a ajutat la decodificarea mesajelor ce fuseseră criptate de mașina de codificat *Enigma* a naziștilor. În anul 1943, o echipă de ingineri influențată de Turing a terminat ceea ce se poate demonstra că a fost primul computer, *Colossus*, care le-a permis Aliaților să continue decodificarea mesajelor criptate cu o versiune mai sofisticată a *Enigmei*. *Bombe* și *Colossus* erau

concepute pentru o sarcină unică şi nu puteau fi reprogramate pentru o alta. Dar se achitau de această sarcină într-un mod strălucit şi sunt creditate cu faptul că le-au permis Aliaților să depăşească avantajul de trei la unu de care se bucura aviația germană Luftwaffe în fața Forțelor Aeriene Regale britanice şi să câştige cruciala Bătălie a Britaniei, dar şi să continue să anticipeze tactica naziştilor pe tot parcursul războiului.

Pe acest fundament a creat John Von Neumann arhitectura computerului modern, care reprezintă cea de-a treia idee majoră a noastră. Denumită „maşina Neumann", ea a rămas practic structura centrală a oricărui computer din ultimii şaizeci şi şapte de ani, de la microcontrolul din maşina de spălat, la cele mai mari supercomputere. Într-o lucrare datată 30 iunie 1945 şi intitulată *Prima schiță a unui raport despre EDVAC*, von Neumann a prezentat ideile care au dominat computația de atunci încoace.[3] Modelul von Neumann include o unitate centrală de procesare, în care se realizează operațiile aritmetice şi logice; o unitate de memorie, în care sunt stocate programul şi datele; memoria de masă; un contor de program şi canalele de input/output. Deşi această lucrare era destinată să fie un document intern al proiectului, a devenit Biblia creatorilor de computere. Nu poți şti când se va întâmpla ca o memorie internă de rutină să pună capăt revoluționării lumii.

Nu s-a intenționat ca maşina Turing să fie practică. Teoremele lui Turing s-au ocupat nu de eficiența rezolvării problemelor în examinarea gamei de probleme care puteau fi rezolvate, teoretic, de computație. Scopul lui von Neumann, pe de altă parte, a fost acela de a crea un concept fezabil de maşină de computat. Modelul lui înlocuieşte computațiile bit-cu-bit ale lui Turing cu versiunea bit-multiple (în general, multipli de opt biți). Banda de memorie a lui Turing este secvențială, astfel încât programele pentru maşina Turing consumă excesiv de mult timp pentru mutarea benzii înainte şi înapoi în scopul stocării şi recuperării rezultatelor intermediare. Dimpotrivă, memoria lui von Neumann are acces aleatoriu, astfel încât orice element de date poate fi recuperat imediat.

Una dintre ideile-cheie ale lui von Neumann a fost programul stocat, pe care îl introdusese cu un deceniu înainte: plasarea programului în acelaşi tip de memorie cu acces aleatoriu ca şi datele (şi adeseori în acelaşi bloc de memorie). Acest fapt permite ca un computer să fie reprogramat atât pentru diferite funcții, cât şi pentru codul de automodificare (dacă stocarea programului este inscriptibilă), operație ce facilitează o formă eficientă de recursivitate. Până atunci, practic toate computerele, inclusiv *Colossus*, fuseseră

construite pentru o funcție specifică. Programul stocat oferă unui computer posibilitatea de a fi cu adevărat universal, adeverind prin asta viziunea lui Turing privitoare la universalitatea computației.

Un alt aspect-cheie al mașinii von Neumann este acela că fiecare instrucțiune include un cod ce specifică operația aritmetică sau logică pe care mașina urmează să o realizeze și adresa unui operant din memorie.

Conceptul lui von Neumann cu privire la modul cum trebuie proiectat un computer a fost lansat odată cu publicarea proiectului EDVAC, realizat împreună cu colaboratorii J. Presper Eckert și John Mauchly. EDVAC nu a funcționat propriu-zis până în anul 1951, moment în care existau și alte computere cu programe stocate, precum Manchester Small-Scale Experimental Machine, ENIAC, EDSAC și BINAC, toate profund influențate de lucrarea lui von Neumann și avându-i ca proiectanți pe Eckert și Mauchly. Von Neumann a contribuit direct la conceperea mai multor mașini de acest fel, inclusiv a unei versiuni ulterioare a lui ENIAC, care includea un program stocat.

Au existat câțiva precursori ai arhitecturii lui von Neumann, deși, cu o excepție surprinzătoare, nici măcar unul nu a fost o mașină von Neumann veritabilă. În anul 1944, Howard Aiken a lansat Mark I, care avea un element de programabilitate, dar nu folosea un program stocat. Citea instrucțiuni de pe o bandă de hârtie perforată și apoi executa imediat comanda, însă îi lipsea o instrucțiune pe ramură condițională.

În anul 1941, omul de știință german Konrad Zuse (1910–1995) a creat computerul Z-3. Și acesta putea să-și citească programul de pe o bandă (în cazul de față, codificat pe film), dar nu avea nici el instrucțiuni pe ramură condițională. Fapt interesant, Zuse s-a bucurat de ajutor din partea Institutului German pentru Cercetări Aeronautice, care a folosit acest dispozitiv pentru studierea vibrațiilor aripilor de avion, dar propunerea făcută de el guvernului nazist, de finanțare a înlocuirii releelor sale cu tuburi vidate, a fost respinsă. Naziștii au considerat că acest domeniu al computației nu era „important pentru război". Această perspectivă contribuie foarte mult, din punctul meu de vedere, la explicarea rezultatului războiului.

De fapt, există un precursor veritabil al conceptului lui von Neumann și acesta vine de dinainte de el cu un secol! Motorul Analitic al matematicianului și inventatorului englez Charles Babbage (1791-1871), care a fost descris pentru prima oară în anul 1837, încorpora ideile lui von Neumann și avea un program stocat pe cartoane perforate, împrumutate de la războiul de țesut Jacquard.[4] Memoria cu acces aleatoriu a Motorului includea 1.000

de cuvinte de 50 de cifre zecimale fiecare (echivalentul a 21 de kilobiți). Fiecare instrucțiune includea un cod de operare și un număr operativ, exact ca limbajele pentru mașinile moderne. Includea și ramificarea condițională și buclele, astfel încât era o mașină von Neumann veritabilă. Motorul Analitic era exclusiv mecanic și se pare că depășea abilitățile de concepție și organizare ale lui Babbage, care a construit componente ale acestei mașini, dar ea nu a funcționat niciodată. Este neclar dacă pionierii computerului din secolul XX, inclusiv von Neumann, erau la curent cu cercetările lui Babbage.

Computerul lui Babbage s-a concretizat într-adevăr în crearea domeniului numit programare software. Scriitoarea engleză Ada Byron (1815-1852), contesă de Lovelace și singurul copil legitim al poetului Lord Byron, a fost primul programator de computer din istorie. Ea a scris pentru Motorul Analitic programe pe care trebuia să le corecteze în minte (de vreme ce computerul nu a funcționat niciodată), practică foarte bine cunoscută inginerilor software din zilele noastre sub denumirea de „verificare la masă". Lovelace a tradus un articol al matematicianului italian Luigi Menabrea pe tema Motorului Analitic și a adăugat textului note proprii cuprinzătoare, scriind că „Motorul Analitic țese modele algebrice exact la fel cum războiul de țesut Jacquard țese flori și frunze". Ada Byron a mers și mai departe și a făcut probabil primele speculații cu privire la fezabilitatea inteligenței artificiale, dar a concluzionat că Motorul Analitic nu are „niciun fel de pretenții de a da naștere la ceva".

Concepția lui Babbage este absolut miraculoasă dacă e să ții seama de epoca în care a trăit și a lucrat el. În orice caz, până la mijlocul secolului XX, ideile lui se pierduseră în negura timpului (cu toate că au fost redescoperite ulterior). Von Neumann are meritul de a fi cel care a conceptualizat și articulat principiile-cheie ale computerului așa cum îl cunoaștem noi astăzi, iar lumea recunoaște acest fapt continuând să se refere la mașina von Neumann ca la principalul model de computație. Reține, totuși, cititorule, că mașina von Neumann comunică în permanență date între diferitele sale unități și în cadrul acestor unități, astfel încât nu putea fi construită fără teoremele lui Shannon și metodele pe care le-a conceput acesta pentru transmiterea și stocarea informației digitale sigure.

Iar asta ne conduce la cea de-a patra idee importantă, care ne invită să trecem dincolo de concluzia Adei Byron că un computer nu poate gândi în mod creator și nu poate găsi algoritmi-cheie folosiți de creier și să purcedem la transformarea computerului într-un creier. Alan Turing a lansat acest

obiectiv în lucrarea sa din anul 1950 *Computing Machinery and Intelligence – Maşini şi inteligenţă computabile*, care include actualmente celebrul său Test Turing pentru stabilirea certă a faptului că o inteligenţă artificială a atins nivelul uman de inteligenţă.

În anul 1956, von Neumann a început să pregătească o serie de prelegeri destinate prestigioasei serii Silliman de la Universitatea Yale. Din cauza ravagiilor cancerului, savantul nu a mai susţinut aceste prelegeri şi nici nu a terminat manuscrisul de pe care urmau să fie citite ele. Acest document neterminat rămâne totuşi un precursor strălucit şi profetic al proiectului pe care eu îl consider a fi cel mai îndrăzneţ şi mai important al omenirii. A fost publicat postum, sub titlul *The Computer and the Brain – Computerul şi creierul,* în anul 1958. Este cât se poate de potrivit ca ultima lucrare a unuia dintre cei mai străluciţi matematicieni ai secolului XX, unul dintre pionierii erei computerelor, să fie întruchiparea inteligenţei. Acest proiect a reprezentat cea mai serioasă cercetare timpurie a creierului uman din perspectiva matematicianului şi a specialistului în computere. Înaintea lui von Neumann, domeniile ştiinţei computerelor şi neuroştiinţei erau două insule fără vreo punte între ele.

Von Neumann îşi începe discuţia scoţând în evidenţă asemănările şi deosebirile dintre computer şi creierul uman. Dat fiind momentul în care a fost elaborat, acest manuscris este remarcabil de exact. Von Neumann a notat faptul că output-ul neuronilor este digital – axonul fie emite, fie nu emite. Acest lucru era departe de a fi evident la vremea respectivă. Însă procesarea la nivelul dendritelor care duc la neuron şi în corpul propriu-zis al neuronului s-a dovedit a fi analogică, iar savantul şi-a prezentat calculele sub forma însumării greutăţii imput-urilor cu un prag. Acest model de funcţionare a neuronului a condus la domeniul conectivităţii, care a construit sisteme bazate pe respectivul model atât ca hardware, cât şi ca software. (După cum am arătat în capitolul anterior, primul asemenea sistem a fost creat de Frank Rosenblatt ca program software pe un computer IBM 704 la Universitatea Cornell, în anul 1957, imediat după ce au devenit accesibile prelegerile lui von Neumann.) Actualmente avem modele mai sofisticate ale felului în care neuronii combină input-urile, dar ideea fundamentală a procesării analogice a input-ului dendritelor folosind concentraţii ale neurotransmiţătorilor a rămas validă.

Von Neumann a aplicat conceptul universalităţii computaţiei pentru a concluziona că, deşi arhitectura şi cărămizile fundamentale par să fie radical

diferite în raportul creier-computer, putem totuşi concluziona că o maşină von Neumann poate simula procesarea la nivelul creierului.

Reciproca nu este însă valabilă, deoarece creierul nu e o maşină von Neumann şi nu are un program stocat ca atare (deşi în mintea noastră putem simula o maşină Turing foarte simplă). Algoritmul şi metodele sunt implicite în structura acesteia. Von Neumann concluzionează corect că neuronii pot învăţa din input-uri modele, despre care am stabilit deja că sunt codificate parţial în tăria dendritelor. Ceea ce nu se ştia pe vremea lui von Neumann era faptul că învăţarea are loc şi prin crearea şi distrugerea conexiunilor dintre neuroni.

Von Neumann sublinia, anticipând, că viteza procesării neuronale este extrem de redusă, de ordinul sutelor de calcule pe secundă, dar că creierul compensează acest neajuns prin procesarea masiv paralelă – o altă intuiţie, deloc evidentă, dar cheie. Von Neumann a susţinut că toţi cei 10^{10} neuroni ai creierului (o apreciere destul de precisă; estimările actuale sunt cuprinse între 10^{10} şi 10^{11}) procesează în acelaşi timp cu ceilalţi. De fapt, toate conexiunile (cu o medie de 10^3 până la 10^4 conexiuni pentru un neuron) computează simultan.

Von Neumann şi descrierile pe care le face el procesării neuronale sunt remarcabile, dată fiind starea de început a neuroştiinţei la momentul respectiv. Un aspect al lucrării sale cu care nu sunt de acord îl reprezintă evaluarea făcută de el în legătură cu capacitatea memoriei creierului. Savantul presupune că creierul îşi aminteşte fiecare input pentru tot restul vieţii. Von Neumann ia în calcul o durată medie de viaţă de 60 de ani sau de aproximativ 2×10^9 secunde. La aproximativ 14 input-uri pentru fiecare neuron în fiecare secundă (ceea ce înseamnă de fapt cu cel puţin trei ordine de mărime mai puţin) şi la 10^{10} neuroni, el ajunge la o estimare de circa 10^{20} de biţi pentru întreaga capacitate a memoriei creierului. Realitatea, după cum am subliniat anterior, este că ne amintim numai un eşantion foarte redus din gândurile şi experienţele noastre şi că amintirile nu sunt stocate ca modele în biţi la un nivel inferior (cum ar fi o imagine video), ci mai curând ca succesiuni de modele de nivel superior.

Descriind fiecare mecanism al creierului, von Neumann arată că un computer modern ar putea face acelaşi lucru în ciuda deosebirilor aparente. Mecanismele analogice ale creierului pot fi simulate prin cele digitale, deoarece computaţia digitală poate emula valori analogice la orice grad de precizie (iar precizia informaţiei analogice din creier este relativ redusă).

Paralelismul masiv al creierului poate fi simulat şi el, dat fiind avantajul semnificativ al vitezei computerelor în computația în serie (un avantaj care s-a extins foarte mult de-a lungul timpului). În plus, putem folosi procesarea în paralel la computere prin utilizarea maşinilor paralele von Neumann – exact cum funcționează computerele din zilele noastre.

Von Neumann concluzionează că metodele creierului nu pot implica algoritmi secvențiali lungi, dacă se ia în considerare rapiditatea cu care pot oamenii să ia decizii, combinată cu viteza foarte redusă de computare a neuronilor. Când jucătorul de la baza a treia are o minge în joc şi decide să o arunce spre prima bază şi nu spre a doua, ia această decizie într-o fracțiune de secundă, adică un timp suficient numai pentru ca fiecare neuron să treacă prin câteva cicluri. Von Neumann concluzionează corect că puterile remarcabile ale creierului sunt conferite de faptul că toate cele 100 de miliarde de neuroni sunt capabile să proceseze informația simultan. După cum am subliniat, cortexul vizual face judecăți vizuale sofisticate în numai trei sau patru cicluri neuronale.

Creierul are o plasticitate considerabilă, care ne permite să învățăm. Însă computerul are o plasticitate şi mai mare, care poate să-i restructureze complet metodele, prin modificarea software-ului său. Prin urmare, sub acest aspect, computerul va putea să imite creierul, însă reciproca nu este valabilă.

Când von Neumann a comparat capacitatea organizării masiv paralele a creierului cu cea a computerelor vremii (cele câteva), a devenit clar că creierul avea o capacitate mult mai mare de memorie şi viteză. A fost construit deja primul supercomputer care a îndeplinit specificațiile corespunzătoare unora dintre cele mai conservatoare estimări ale vitezei necesare pentru simularea funcțională a creierului uman (în jur de 10^{16} operații pe secundă).[5] (Personal, estimez că acest nivel de computare va costa 1.000 de dolari până la începutul anilor 2020.) În privința memoriei, suntem chiar şi mai aproape. Deşi şi-a redactat acest remarcabil manuscris devreme în cadrul istoriei computerului, von Neumann a avut totuşi încredere că atât hardware-ul, cât şi software-ul inteligenței umane îşi vor afla în cele din urmă locul, aceasta fiind şi motivația lui pentru pregătirea prelegerilor despre care am pomenit.

Von Neumann a fost cât se poate de conştient de ritmul tot mai accelerat al progresului şi de implicațiile profunde ale acestuia în viitorul omenirii. La un an după moartea sa, în anul 1957, confratele matematician Stan Ulam l-a citat că ar fi spus, la începutul anilor 1950, că „progresul din ce în ce

mai accelerat al tehnologiei şi schimbările modului de viaţă al omului dau aparenţa apropierii unei singularităţi esenţiale în istoria rasei, dincolo de care treburile omului, aşa cum le cunoaştem, nu vor mai putea continua". Aceasta este prima utilizare cunoscută a cuvântului „singularitate" în contextul istoriei tehnologice umane.

Ceea ce a intuit în mod fundamental von Neumann a fost faptul că există o echivalenţă esenţială între computer şi creier. A se observa că inteligenţa emoţională a omului biologic este parte a inteligenţei sale generale. Dacă intuiţia lui von Neumann este corectă şi dacă se acceptă ideea mea că o entitate nonbiologică în stare să recreeze inteligenţa (emoţională şi de alte genuri) a unui om biologic este conştientă (a se vedea capitolul următor), atunci ar trebui să se ajungă la concluzia că există o echivalenţă esenţială între computer – cu software-ul adecvat – şi minte (mintea conştientă). Prin urmare, von Neumann a avut dreptate?

Majoritatea computerelor din zilele noastre sunt exclusiv digitale, în vreme ce creierul omului combină metodele digitală şi analogică. Dar metodele analogice sunt uşor şi rutinier recreate de cele digitale la orice nivel de acurateţe. Specialistul american în ştiinţa computerelor Carver Mead (născut în 1934) a arătat că putem să imităm în siliciu în mod direct metodele analogice ale creierului, fapt demonstrat cu ceea ce el numea cipuri „neuromorfice".[6] Mead a arătat cum este posibil ca această metodă să fie de mii de ori mai eficientă decât metodele digitale care le imită pe cele analogice. Cu cât codificăm algoritmul masiv recurent al neocortexului, cu atât devine mai logică folosirea abordării lui Mead. Grupul de Computare Cognitivă de la IBM, condus de Dharmendra Modha, a lansat cipuri care imită neuronii şi conexiunile lor, având inclusiv capacitatea de a stabili noi conexiuni.[7] Denumit SYNAPSE, unul dintre cipuri asigură o simulare directă a unui număr de 256 de neuroni cu aproximativ un sfert de milion de conexiuni sinaptice. Scopul proiectului este crearea unui neocortex simulat cu 10 miliarde de neuroni şi 100 de mii de miliarde de conexiuni – aproape cât un creier uman –, care foloseşte doar un kilowatt de putere.

Aşa cum a spus von Neumann în urmă cu peste o jumătate de secol, creierul este extrem de lent, dar masiv paralel. Circuitele digitale din zilele noastre sunt de cel puţin 10 milioane de ori mai rapide decât comutatoarele electrochimice ale creierului. Reciproc, toţi cei 300 de recognizatori de modele din neocortexul creierului procesează simultan şi întregul număr de o sută de mii de miliarde de conexiuni interneuronale computează concomitent.

Însă aspectul-cheie în ceea ce privește asigurarea hardware-ului necesar pentru realizarea cu succes a modelului unui creier uman îl reprezintă memoria de ansamblu și capacitatea de calcul impusă. Noi nu trebuie să copiem direct arhitectura creierului, aceasta fiind o abordare foarte ineficientă și inflexibilă.

Hai să estimăm care ar fi aceste cerințe hardware. Multe proiecte au încercat să imite tipul ierarhic de învățare și de recunoaștere a modelelor care se realizează la nivelul neocortexului, inclusiv cercetările mele cu Modele Markov Ierarhice Ascunse. O estimare conservatoare extasă din propria mea experiență spune că imitarea unui ciclu într-un singur recognizator de modele din neocortexul biologic al creierului ar avea nevoie de un număr de circa 3.000 de calcule. Majoritatea simulărilor se realizează la o jumătate din această estimare. Când creierul funcționează cu circa 10^2 (100) de cicli pe secundă, ajungem la 3×10^5 (300.000) de calcule pe secundă pe recognizator de modele. Folosind estimarea mea de 3×10^8 (300 de milioane) de recognizatori de modele, obținem în jur de 10^{14} (100 de mii de miliarde) de calcule pe secundă, o cifră care corespunde estimării mele din cartea *The Singularity Is Near – Singularitatea este aproape.* În această carte, am anticipat că simularea funcțională a creierului ar impune între 10^{14} și 10^{16} calcule pe secundă (cps) și am folosit 10^{16} calcule pentru a fi conservator. Estimarea lui Hans Moravec, specialist în inteligență artificială, bazată pe extrapolarea cerințelor computaționale ale procesării vizuale timpurii (inițiale) la nivelul întregului creier este de 10^{14} cps, cifră ce corespunde estimării mele.

Mașinile obișnuite de birou pot să ajungă la 10^{10} cps, deși acest nivel de performanță poate fi amplificat semnificativ prin folosirea resurselor din nor. Cel mai rapid supercomputer, K Computer din Japonia, a ajuns deja la 10^{16} cps.[8] Dat fiind faptul că algoritmul neocortexului este masiv repetat, abordarea folosirii de cipuri neuromorfice, așa cum este IBM SYNAPSE, menționat mai înainte, este de asemenea promițătoare.

În ceea ce privește necesarul de memorie, ne trebuie în jur de 30 de biți (aproximativ patru byte) pentru ca o conexiune să se lege de unul din totalul de 300 de milioane de recognizatori de modele. Dacă estimăm o medie de opt input-uri pentru fiecare recognizator, ajungem la 32 de byte pentru unul singur. Dacă adăugăm o greutate de un byte pentru fiecare input, ajungem la 40 de byte. Mai punem 32 de byte pentru conexiunile descendente și ajungem la 72 de byte. Observă, cititorule, că numărul care cuantifică ramificarea în jos și în sus va fi întotdeauna mult mai mare de opt, cu toate că

aceşti arbori foarte ramificaţi sunt împărţiţi între ei de numeroşi recognizatori. De exemplu, pot să existe sute de recognizatori implicaţi în recunoaşterea literei „p". Aceştia transmit semnale câtorva mii de asemenea recognizatori de la următorul nivel, care se ocupă de cuvinte şi expresii ce conţin „p". Pe de altă parte, recognizatorii de „p" nu repetă arborele de conexiuni care deserveşte toate cuvintele şi expresiile ce conţin „p" – toţi folosesc în comun un asemenea arbore de conexiuni. Aceeaşi observaţie este valabilă şi pentru conexiunile descendente: recognizatorul răspunzător pentru cuvântul „APPLE" le va spune miilor de recognizatori ai lui „E" de la un nivel inferior că este de aşteptat un „E", dacă au fost identificate deja „A", „P", „P" şi „L". Acest arbore de conexiuni nu se repetă pentru fiecare recognizator de cuvânt sau de expresie care vrea să informeze următorul nivel inferior că este de aşteptat un „E". Iarăşi, este folosit în comun. Pentru acest motiv, o estimare de ansamblu de opt ascendent şi opt descendent – în medie, desigur – pentru un recognizator de model este ceva rezonabil. Chiar dacă mărim această estimare, ea nu modifică semnificativ ordinul de mărime al estimării rezultante.

Cu 3 × 108 (300 de milioane) de recognizatori de modele la 72 de byte fiecare, obţinem un necesar total de memorie de aproximativ 2×10^{10} (20 de miliarde) de byte. Este o cifră cât se poate de modestă, pe care computerele uzuale din zilele noastre o pot depăşi.

Aceste estimări sunt făcute doar ca să ofere aproximări ale ordinului de mărime cerut. Dat fiind faptul că circuitele digitale sunt în mod inerent de circa 10 milioane de ori mai rapide decât circuitele neocorticale biologice, nu trebuie să egalăm paralelismul creierului uman – o procesare paralelă modestă (comparativ cu paralelismul de miliarde de ori mai mare al creierului uman) va fi suficientă. Se poate vedea că necesităţile computaţionale ajung astfel la îndemână. Recablarea permanentă a creierului – dendritele îşi creează noi sinapse continuu – poate fi de asemenea imitată de software cu ajutorul link-urilor, un sistem mult mai flexibil decât cel care implică plasticitatea creierului, impresionantă dar limitată.

Redundanţa folosită de creier pentru atingerea unor rezultate invariante solide poate fi imitată în mod cert în emulaţiile software. Suportul matematic al optimizării acestor tipuri de sisteme de învăţare ierarhice autoorganizate este bine înţeleasă. Organizarea creierului e departe de a fi optimă. Desigur, nici n-a trebuit să fie – a trebuit să fie doar suficient de bună pentru atingerea pragului de la care creierul să poată fi capabil să creeze instrumente care să-i compenseze propriile limitări.

O altă restricție a neocortexului uman spune că nu există vreun proces care să elimine sau măcar să treacă în revistă ideile contradictorii, fapt ce explică de ce gândirea umană este adeseori masiv inconsecventă. Mecanismul care se ocupă de această așa-numită gândire critică e slab, iar abilitatea nu-i este exercitată nici pe departe atât de des cât ar trebui. Într-un neocortex bazat pe software, putem include un proces care să dezvăluie contradicțiile, pentru a le verifica pe mai departe.

Este important să se observe că alcătuirea unei întregi regiuni a creierului este mai simplă decât compunerea unui singur neuron. Așa cum am discutat anterior, modelele devin adeseori mai simple la un nivel superior – să ne gândim la analogia cu un computer. Trebuie neapărat să înțelegem fizica semiconductorilor pentru a crea modelul unui tranzistor, iar ecuațiile aflate la baza unui singur tranzistor real sunt complexe. Un circuit digital care multiplică două numere are nevoie de sute. Cu toate acestea, putem crea un model al acestui circuit de multiplicare foarte simplu, cu una sau două formule. Pentru un întreg computer, cu miliarde de tranzistori, se poate face un model prin intermediul setului de instrucțiuni și al descrierii registrelor acestuia, descriere ce necesită câteva foi pentru text și formule. Programele software pentru un sistem de operare, compilatoarele de limbaj și asamblatoarele sunt complexe la un nivel rezonabil, însă realizarea modelului unui anumit program – de exemplu, un program de recunoaștere a vorbirii bazat pe Modele Markov Ascunse Ierarhice – poate fi la fel de bine descris pe numai câteva pagini cu ecuații. Nicăieri într-o asemenea descriere nu s-ar putea găsi detalii privitoare la fizica semiconductorilor sau chiar la arhitectura computerului.

O observație similară este adevărată și în cazul creierului. Un recognizator neocortical de modele care detectează o anumită caracteristică vizuală invariabilă (de pildă, un chip), care realizează o filtrare de *bandpass* (restricționarea input-ului la o gamă de frecvențe specifice) a sunetului sau care evaluează proximitatea temporală a două evenimente poate fi descris cu mult mai puține detalii specifice decât relațiile fizice și chimice propriu-zise care controlează neurotransmițătorii, canalele de ioni și alte variabile sinaptice și dendritice implicate în procesele neuronale. Deși această complexitate trebuie să fie atent cântărită înaintea trecerii la următorul nivel conceptual superior, mare parte din ea poate fi simplificată, pe măsură ce sunt dezlegate principiile de funcționare a creierului.

Capitolul 9

EXPERIMENTE MINTALE DESPRE MINTE

Mințile sunt pur şi simplu ceea ce fac creierele.

— Marvin Minsky, Societatea Minții

Când se vor construi maşini inteligente, nu ar trebui să fim surprinşi să constatăm că acestea sunt la fel de confuze şi încăpățânate ca oamenii în convingerile lor referitoare la subiectul minții, al conştienței, al liberului arbitru şi al altora asemenea.

— Marvin Minsky, Societatea Minții

Cine este conştient?

Adevărata istorie a conştienței începe cu prima minciună a cuiva.

— Iosif Brodsky

Suferința este unica origine a conştiinței.

— Feodor Dostoievski, *Însemnări din subterană*

Există un soi de plantă care, cu ajutorul florilor sale, consumă hrană organică: atunci când pe floare se aşază o muscă, petalele i se închid şi țin musca strâns până când planta o absoarbe în organismul ei; dar floarea nu se închide decât la ce e bun de mâncat; unei picături de ploaie sau unui beţişor nu-i acordă vreo atenție. Ciudat ca un lucru atât de lipsit de conştiință să fie atât de atent la propriul interes. Dacă asta înseamnă inconştiență, care e folosul conştienței?

— Samuel Butler, 1871

Am examinat creierul ca entitate capabilă de anumite niveluri de a duce lucrurile la bun sfârşit. Dar această perspectivă, în esență, ne lasă în afara tabloului. Părem să trăim în creierul nostru. Avem vieți subiective. Cum se relaționează concepția noastră obiectivă asupra creierului, pe care am

discutat-o până acum, cu sentimentele noastre, cu sentimentul că suntem persoana care trăiește aceste experiențe?

Filosoful britanic Colin McGinn (născut în anul 1950) scrie că discutarea „conștienței poate să-l reducă până și pe cel mai pretențios gânditor la o incoerență bâlbâită". Motivul este că oamenii au adeseori concepții neverificate și inconsecvente cu privire la semnificațiile acestui termen.

Mulți observatori consideră conștiența ca fiind o formă de performanță – de exemplu, capacitatea de reflectare asupra sinelui, respectiv abilitatea de a-ți înțelege propriile gânduri și de a le explica. Eu aș descrie conștiența ca fiind abilitatea de a gândi asupra propriei gândiri. După câte s-ar putea presupune, am putea găsi o cale oarecare de evaluare a acestei abilități și apoi de a folosi acest test pentru separarea lucrurilor conștiente de lucrurile inconștiente.

Pe de altă parte, însă, intrăm repede la încurcătură dacă încercăm să implementăm această abordare. Sugarul este conștient? Dar câinele? Nici unul și nici celălalt nu sunt prea pricepuți la descrierea propriilor procese de gândire. Există oameni care cred că sugarii și câinii nu sunt ființe conștiente tocmai din cauză că nu se pot autoexplica. Dar cum rămâne atunci cu computerul pe nume Watson? E posibil să fie pus și el într-un mod de funcționare în care să ne poată practic lămuri cum de a dat un anumit răspuns? Deoarece conține un model al propriei gândiri, Watson este, prin urmare, conștient, în vreme ce sugarul și câinele nu sunt?

Înainte de a continua analizarea acestei chestiuni, este important să reflectăm asupra celei mai importante distincții legate de ea: Ce anume putem lămuri de pe urma științei versus ce rămâne cu adevărat o problemă de filosofie? Un punct de vedere este acela că filosofia reprezintă un fel de refugiu pentru întrebări care nu au fost încă supuse metodei științifice. Potrivit acestei perspective, din momentul în care o știință progresează suficient de mult pentru a rezolva un anumit set de chestiuni, filosofii pot trece la alte preocupări, până sosește momentul în care știința le rezolvă și pe ele. Această perspectivă este endemică în ceea ce privește chestiunea conștienței și mai ales întrebarea „Ce și cine este conștient?"

Să ne gândim la aceste afirmații ale filosofului John Searle: „Știm că creierul produce conștiența prin mecanisme biologice specifice... Esențial este să se recunoască faptul că conștiența e un proces biologic, asemenea digestiei, lactației, fotosintezei sau mitozei... Creierul este o mașină – o mașină biologică, pentru a fi mai exact, dar, totuși, o mașină. Prin urmare,

primul pas este să se stabilească felul cum face creierul acest lucru şi apoi să se construiască o maşină artificială care să aibă acelaşi mecanism, la fel de eficient, pentru producerea conştienţei".[1] Oamenii sunt adeseori surprinşi să vadă citate precum acesta, deoarece ei presupun că Searle e devotat protejării misterului conştienţei de către reducţionişti de felul lui Ray Kurzweil.

Filosoful australian David Chalmers (născut în anul 1966) a inventat termenul „problema grea a conştienţei" pentru a descrie dificultatea identificării cu precizie a acestui concept în esenţă indescriptibil. Uneori, o expresie scurtă rezumă o întreagă şcoală de gândire atât de bine, încât devine emblematică (de exemplu, „banalitatea răului", a lui Hannah Arendt). Celebra formulare a lui Chalmers face şi ea foarte bine acest lucru.

Discutând despre conştienţă, este foarte uşor să aluneci către cercetarea atributelor observabile şi măsurabile pe care le asociem cu faptul de a fi conştient, dar acestei abordări îi lipseşte însăşi esenţa ideii. Tocmai am menţionat conceptul de metacogniţie – ideea de a gândi asupra propriei gândiri – ca fiind o asemenea corelare a conştienţei. Alţi observatori combină inteligenţa emoţională sau inteligenţa morală cu conştienţa. Dar, din nou, abilitatea noastră de a exprima un sentiment de iubire, de a pricepe o glumă sau de a fi sexy conţine, pur şi simplu, tipuri de performanţe – impresionante şi inteligente, poate, dar, cu toate acestea, abilităţi ce pot fi observate şi măsurate (chiar dacă mai discutăm asupra felului cum să le evaluăm). Descoperirea modului în care creierul realizează acest gen de sarcini şi a fenomenelor ce se petrec în creier atunci când sunt îndeplinite asemenea sarcini constituie întrebarea „simplă" a lui Chalmers. Desigur, problema „simplă" este, de fapt, oricum altcumva şi reprezintă, poate, cea mai dificilă şi mai importantă ţintă ştiinţifică a epocii noastre. Problema „grea" a lui Chalmers este atât de grea, încât e, mai presus de toate, inexprimabilă.

În sprijinul acestei distincţii, Chalmers introduce un experiment mintal care implică ceea ce el numeşte *zombies*. Zombie este o entitate care se poartă exact ca un om, dar pur şi simplu nu are nicio experienţă subiectivă – cu alte cuvinte, zombie nu are conştienţă. Chalmers susţine că, de vreme ce putem să îi concepem, zombii sunt cel puţin logic posibili. Dacă te-ai afla la un cocteil la care ar fi prezenţi atât oameni „normali", cât şi *zombies*, cum ţi-ai da seama de deosebirile dintre ei? Poate că situaţia se aseamănă cu cocteilul la care chiar ai fost.

Mulţi oameni răspund la această întrebare afirmând că i-ar interoga pe indivizii pe care ar dori să-i evalueze în privinţa reacţiilor lor emoţionale la

evenimente și idei. Un zombie, cred ei, și-ar trăda lipsa de experiență subiectivă printr-o deficiență manifestată în anumite tipuri de răspunsuri emoționale. Dar un răspuns pe aceste coordonate pur și simplu nu conține ipotezele experimentului mintal. Dacă am întâlni o persoană neemoțională (de pildă un individ cu deficiențe emoționale, așa cum se întâmplă în anumite cazuri de autism), un avatar sau un robot care nu este convingător ca ființă umană emoțională, entitatea nu ar fi un zombie. Reține, cititorule: potrivit ipotezei lui Chalmers, un zombie este absolut normal sub aspectul capacității sale de a răspunde, inclusiv sub cel al abilității de a reacționa emoțional; doar că îi lipsește experiența subiectivă. Ideea de bază este că nu există vreun mijloc de identificare a unui zombie, deoarece, prin definiție, în comportament nu există niciun indiciu aparent al naturii de zombie. Deci e vorba despre o distincție fără diferențe?

Chalmers nu încearcă să răspundă la întrebarea grea, dar oferă niște probabilități. Una constă într-o formă de dualism în care, în lumea fizică, nu există conștiență ca atare, ci mai curând ca realitate ontologică separată. Potrivit acestei formulări, ceea ce face un om se bazează pe procesele din creierul lui. Deoarece creierul este închis din punct de vedere cauzal, putem explica pe deplin acțiunile unei persoane, inclusiv gândurile sale, prin intermediul proceselor din creierul său. Pe urmă, conștiența există într-un alt domeniu sau cel puțin este o proprietate diferită de a proceselor din lumea fizică. Această explicație nu permite minții (cu alte cuvinte, proprietății conștiente asociate creierului) să afecteze creierul în mod cauzal.

O altă probabilitate pe care o enunță Chalmers și care nu este distinctă sub aspect logic de concepția lui despre dualism, fiind adeseori numită panprotopsihism, susține că toate sistemele fizice sunt conștiente, deși omul este mai conștient decât, să spunem, un întrerupător. Cu siguranță, aș fi de acord că creierul omului are mult mai multe lucruri de care să fie conștient decât un întrerupător de lumină.

Concepția mea, care e probabil o subșcoală a panprotopsihismului, sună așa: conștiența e o proprietate emergentă a unui sistem fizic complex. Din această perspectivă, câinele este și el conștient, dar pe undeva mai puțin decât omul. Și furnica are un anumit nivel de conștiență, însă mai redus decât al câinelui. Colonia de furnici, pe de altă parte, se poate considera că are un nivel de conștiență mai înalt decât furnica individuală; adică, în mod cert, colonia e mai conștientă decât o singură furnică. Potrivit acestei socoteli, un computer care imită cu succes complexitatea creierului uman ar avea aceeași conștiență emergentă ca un om.

Un alt mod de a conceptualiza ideea de conştienţă o imaginează pe aceasta ca sistem care are „qualia". Dar ce înseamnă qualia? O definiţie a termenului este „experienţe conştiente". Însă definiţia asta nu ne duce prea departe. Gândeşte-te la următorul experiment mintal: un specialist în neuroştiinţă este complet orb la culori – nu genul de orbire din cauza căreia cineva amestecă anumite culori, să spunem verde şi roşu (ca mine), ci, dimpotrivă, o afecţiune din cauza căreia individul afectat trăieşte într-o lume exclusiv alb-negru. (Într-o versiune extremă a acestui scenariu, persoana respectivă a crescut într-o lume în alb-negru şi nu a văzut niciodată vreo culoare. Ideea este că, în lumea persoanei respective, nu există culori.) Însă persoana cu pricina a studiat amănunţit fizica cromatică – este conştientă că lungimea de undă a luminii roşii e de 700 nanometri – şi procesele neurologice ale unei persoane care trăieşte în mod normal experienţa culorilor; prin urmare, ştie mult mai multe despre felul în care creierul procesează culorile. Ştie despre culori mai multe decât cei mai mulţi oameni. Dacă ai vrea să ajuţi respectiva persoană încercând să-i explici cum este de fapt această trăire a experienţei „roşii", cum ai face-o? Poate că ar trebui să-i citeşti o parte din poezia *Roşu* a poetului nigerian Oluseyi Oluseun:

Roşu-i culoarea sângelui
simbol al vieţii
Roşu-i culoarea pericolului
simbol al morţii

Roşu-i culoarea trandafirilor
simbol al frumuseţii
Roşu-i culoarea iubiţilor
simbol al unităţii

Roşu-i culoarea roşiei
simbol al sănătăţii
Roşu-i culoarea focului
simbol al dorinţei arzânde

Asta i-ar crea o idee destul de bună despre unele dintre asocierile pe care le fac oamenii cu culoarea roşie şi ar putea chiar să-i faciliteze persoanei susţinerea propriilor asocieri într-o conversaţie despre culori. („Da,

îmi place la nebunie culoarea roşie, e atât de arzătoare şi de aprinsă, atât de periculos de frumoasă...") Dacă ar vrea, probabil că ar putea să convingă pe cineva că a trăit experiența culorii roşii, însă nici toată poezia de pe lume nu i-ar permite să trăiască într-adevăr această experiență.

În mod similar, cum i-ai explica senzația de scufundare în apă cuiva care nu a atins apa niciodată? Am fi iarăşi forțați să recurgem la poezie, dar realmente nu există nicio modalitate de comunicare a experienței propriu-zise. Aceste experiențe constituie ceea ce numim *qualia.*

Mulți dintre cititorii acestei cărți au trăit experiența culorii roşii. Dar de unde pot eu să ştiu dacă experiența culorii roşii pe care ai trăit-o tu nu este aceeaşi cu experiența pe care o trăiesc eu când mă uit la culoarea albastră? Ne uităm amândoi la un obiect roşu şi afirmăm cu certitudine că este roşu, dar asta nu răspunde la întrebarea noastră. E posibil ca eu să am aceeaşi experiență pe care o ai tu când te uiți la culoarea albastră, dar am învățat cu toții să spunem că roşul e roşu. Am putea începe să apelăm iar la poezii, dar ele nu fac altceva decât să reflecte asocierile pe care le realizează oamenii pentru diverse culori; nu vorbesc despre natura de fapt a qualiei. Pe bună dreptate, oamenii orbi congenital au citit o mulțime de lucruri despre culori, asemenea referiri fiind extrem de abundente în literatură şi, prin urmare, acei oameni au o anumită versiune a unei experiențe a culorii. Cum se compară experiența trăită de ei pentru culoarea roşu cu experiența trăită de văzători? Pe bună dreptate, este aceeaşi întrebare ca în cazul femeii care trăieşte într-o lume alb-negru. E remarcabil faptul că fenomene atât de comune din viața noastră sunt complet inexprimabile şi fac imposibilă o confirmare simplă, cum ar fi aceea că trăim experiența aceleiaşi *qualia.*

O altă definiție pentru *qualia* o echivalează pe aceasta cu senzația dată de o experiență. Însă definiția nu e cu nimic mai puțin circulară decât încercările noastre de mai devreme de definire a conştienței: expresiile „Îmi dă senzația", „Trăiesc o experiență" şi „Conştiență" sunt toate sinonime. Conştiența şi chestiunea strâns legată de *qualia* reprezintă o problemă fundamentală, poate întrebarea filosofică supremă (deşi subiectul identității poate să fie chiar mai important, după cum voi arăta în secțiunea de încheiere a acestui capitol).

Deci, în privința conştienței, care este întrebarea? Aceasta: Cine sau ce este conştient? În titlul acestei cărți, mă refer la „minte" şi nu la „creier", deoarece mintea este un creier conştient. Am putea spune de asemenea că mintea are liber arbitru şi identitate. Afirmația că aceste subiecte sunt

filosofice nu este axiomatică. Susțin că ele nu pot fi niciodată pe deplin decise prin intermediul științei. Cu alte cuvinte, nu există experimente falsificabile pe care să le putem avea în vedere pentru a rezolva problema fără a formula ipoteze filosofice. Dacă am construi un detector de conștiență, Searle ar vrea ca dispozitivul să-l asigure că e plin de neurotransmițători biologici. Filosoful american Daniel Dennett (născut în 1942) ar fi mai flexibil în privința substratului, dar ar putea dori să stabilească dacă sistemul conține sau nu un model al lui însuși și al propriei performanțe. Această concepție se apropie mai mult de a mea, dar în esența ei se află în continuare o ipoteză filosofică.

Au fost avansate în permanență propuneri de teorii științifice care leagă conștiența de câte un atribut fizic măsurabil. La acesta Searle se referă denumindu-l „mecanism pentru crearea conștienței". Omul de știință, filosoful și anesteziologul american Stuart Hameroff (născut în anul 1947) scria că „filamentele citoscheletale sunt rădăcinile conștienței".[2] El se referă la firișoarele fine din fiecare celulă (inclusiv neuroni, dar nu numai ei) denumite minitubuli care conferă fiecărei celule integritate structurală și joacă un rol în diviziunea celulară. Cărțile și lucrările lui pe această temă conțin descrieri detaliate și ecuații ce explică plauzibilitatea ca microtubulii să joace un rol în procesarea informației la nivelul celulelor. Dar legătura microtubulilor cu conștiența necesită un salt al credinței care nu este fundamental diferit de saltul credinței în cazul unei doctrine religioase ce descrie o ființă supremă care depozitează conștiența (uneori denumită și „suflet") în anumite entități (de obicei umane). Se aduc unele dovezi slabe în sprijinul concepției lui Hameroff, în speță fiind vorba despre observația că procesele neurobiologice care pot să susțină această presupusă computație celulară sunt stopate în timpul anesteziei. Dar asta e departe de a fi o demonstrație convingătoare, dat fiind faptul că o mulțime de procese sunt oprite în timpul anesteziei. Nu putem spune cu certitudine nici măcar că subiecții nu sunt conștienți sub anestezie. Tot ceea ce știm e că oamenii nu-și amintesc, pe urmă, experiențele. Nici măcar acest lucru nu este universal valabil, deoarece unele persoane pot rememora – cu precizie – experiențele din timpul anesteziei, inclusiv, de exemplu, conversațiile chirurgilor. Denumit conștientizare sub anestezie, acest fenomen apare în Statele Unite de aproximativ de 40.000 de ori pe an.[3] Însă chiar și lăsând asta la o parte, conștiența și memoria sunt concepte complet diferite. După cum am arătat pe larg, dacă mă gândesc la experiențele

mele de peste zi minut-cu-minut, constat că am avut o mulțime de impresii senzoriale, dar îmi amintesc de doar foarte puține dintre ele. Nu am fost conștient de ce am văzut sau auzit toată ziua? E o întrebare bună, dar răspunsul nu este prea clar.

Fizicianul și matematicianul englez Roger Penrose (născut în anul 1931) a făcut un alt salt de credință în propunerea sursei conștienței, cu toate că și teoria lui avea legătură cu microtubulii – mai exact, cu capacitățile lor de computare cuantică. Raționamentul lui Penrose, deși nu a fost afirmat explicit, a părut a fi acela că, la fel ca un efect cuantic, conștiența este misterioasă, astfel încât cele două trebuie să fie legate într-un fel oarecare.

Penrose și-a început analiza cu teoremele lui Turing referitoare la problemele de nerezolvat și cu Teorema incompletitudinii a lui Gödel. Premisa lui Turing (care a fost discutată foarte detaliat în Capitolul 8) spune că există probleme algoritmice care pot fi formulate, dar care nu pot fi rezolvate de o mașină Turing. Dată fiind universalitatea computațională a mașinii Turing, putem concluziona că aceste „probleme insolvabile" nu pot fi rezolvate de nicio mașină. Teorema incompletitudinii a lui Gödel ajunge la un rezultat foarte asemănător în privința abilității de a demonstra ipoteze ce implică numere. Argumentul lui Penrose susține că creierul uman este capabil să rezolve aceste probleme de nerezolvat, deci e capabil și să facă lucruri pe care o mașină deterministă, cum ar fi un computer, nu le poate face. Ideea ce derivă din argumentul lui Penrose este, măcar parțial, aceea că ființa umană se află deasupra mașinilor. Dar premisa centrală – că omul poate rezolva problemele de nerezolvat ale lui Turing și Gödel – este pur și simplu neadevărată, din păcate.

O celebră problemă de nerezolvat, numită problema *busy beaver,* este formulată după cum urmează: Găsește numărul maxim de 1 pe care îl poate scrie pe banda ei o mașină Turing cu un anumit număr de stări. Deci, pentru determinarea valorii *busy beaver* pentru numărul *n*, construim toate mașinile Turing care au *n* stări (un număr finit dacă *n* este finit) și apoi determinăm cel mai mare număr de 1 pe care aceste mașini l-ar putea scrie pe benzile lor, excluzând mașinile Turing care intră într-o buclă infinită. Problema este de nerezolvat deoarece, atunci când încercăm să simulăm toate aceste mașini Turing cu *n* stări, simulatorul nostru va intra într-o buclă infinită în clipa în care va începe să simuleze una dintre mașinile Turing care intră într-o buclă infinită. Pe de altă parte însă, se dovedește că există computere

care au reuşit totuşi să determine funcţia *busy beaver* pentru anumite valori ale lui *n*. La fel şi oamenii, dar computerele au rezolvat problema pentru mult mai multe *n*-uri decât oamenii neasistaţi. Computerele sunt în general mai bune decât omul la rezolvarea problemelor de nerezolvat ale lui Turing şi Gödel.

Penrose a legat aceste presupuse capacităţi transcendentale ale creierului uman de computarea cuantică din interiorul lui. Potrivit lui Penrose, aceste efecte cuantice neuronale sunt cumva inerent inaccesibile computerelor, prin urmare gândirea umană are un avantaj inerent. De fapt, electronica obişnuită foloseşte efecte cuantice (tranzistorii se bazează pe *tunneling*-ul cuantic al electronilor prin bariere); computarea cuantică la nivelul creierului nu a fost demonstrată niciodată; performanţa mintală a omului poate fi explicată satisfăcător cu ajutorul metodelor clasice de computare; şi, în orice caz, nimic nu ne împiedică să aplicăm computarea cuantică la computere. Nici măcar una dintre aceste obiecţii nu i-a fost adresată vreodată lui Penrose. Abia când criticii au subliniat că creierul este un loc prea cald şi dezordonat pentru computarea cuantică, Hameroff şi Penrose şi-au unit forţele. Penrose a găsit în neuroni un vehicul despre care se poate imagina că ar putea să susţină computarea cuantică – şi anume microtubulii, legat de care Hameroff speculase că fac parte din procesarea informaţiei la nivelul neuronului. Aşadar, teza Hameroff-Penrose afirmă că microtubulii din neuroni realizează computarea cuantică şi că această computare este răspunzătoare de conştienţă.

Teza celor doi a fost şi ea criticată, de exemplu de fizicianul şi cosmologul american de origine suedeză Max Tegmark (născut în anul 1967), care a determinat faptul că evenimentele cuantice din microtubuli pot supravieţui doar 10-13 microsecunde, ceea ce înseamnă o perioadă de timp prea scurtă şi pentru computarea unor rezultate cât de cât semnificative, şi pentru afectarea proceselor neuronale. Există anumite tipuri de probleme în cazul cărora computarea cuantică ar dovedi capacităţi superioare în raport cu computarea clasică – de exemplu spargerea codurilor prin factorizarea unor numere mari. Pe de altă parte, gândirea umană neasistată se dovedeşte a fi îngrozitoare în rezolvarea acestor probleme şi nu poate egala nici măcar computerele clasice în acest domeniu, fapt ce sugerează că creierul nu demonstrează niciun fel de capacităţi de computare cuantică. Mai mult de-atât, chiar dacă un fenomen cum este computarea cuantică la nivelul creierului ar exista cu adevărat, el nu ar fi legat de conştienţă în mod obligatoriu.

Trebuie să ai credință

> Ce lucru desăvârșit e omul! Ce aleasă îi e cugetarea! Și cât de nemăsurate însușirile! Ce falnic și minunat e în portul și mișcările lui! Întocmai ca un înger în fapte! Și ca un zeu în năzuințe! Splendoarea lumii! Pilda perfecțiunii vietăților! Și, totuși, ce e pentru mine întruchiparea asta din lut?
>
> — Hamlet, în *Hamlet* de Shakespeare

Realitatea este că aceste teorii sunt toate salturi ale credinței. Aș adăuga că, în ceea ce privește conștiența, principiul călăuzitor spune că „trebuie să ai credință" – cu alte cuvinte, fiecare dintre noi are nevoie de un salt al credinței în privința a cine și ce este conștient și a ce și cine suntem noi ca ființe conștiente. Altminteri, nu ne-am putea trezi dimineața. Dar ar trebui să fim cinstiți cu privire la necesitatea fundamentală a unui salt al credinței în această chestiune și autoreflexivi în privința a ceea ce implică acest salt deosebit al nostru.

Oamenii fac salturi foarte diferite, în ciuda impresiilor de sens contrar. Premisele filosofice individuale privitoare la natura și sursa conștienței stau la baza dezacordurilor pe subiecte care merg de la drepturile animalelor la avort și se concretizează în viitoare conflicte chiar și mai litigioase pe subiectul drepturilor mașinilor. Predicția mea obiectivă este că mașinile din viitor vor părea a fi conștiente și că oamenilor biologici li se vor părea convingătoare atunci când vor vorbi despre percepțiile lor în condițiile experiențelor subiective, conștiente.* Mașinile vor prezenta întreaga gamă a indiciilor subtile, familiar emoționale: ne vor face să râdem și să plângem și se vor înfuria pe noi dacă le vom spune că nu credem că sunt conștiente. (Vor fi foarte inteligente, așa că nu vom dori să se întâmple așa ceva.) Vom ajunge să acceptăm că sunt persoane conștiente. Saltul meu personal de credință este următorul: din momentul în care mașinile vor reuși să fie convingătoare atunci când se vor referi la percepțiile lor în condițiile experiențelor subiective, conștiente, vor fi realmente persoane conștiente. Am ajuns la această poziție prin următorul experiment mintal: imaginează-ți că te întâlnești cu o

---

* [În original, *qualia*] Termen folosit în filosofie cu sensul de instanțe individuale ale experienței subiective, conștiente. Termenul provine din latinescul *qualia*, care înseamnă „ce fel" ori „ce gen". Exemple de qualia: suferința provocată de o durere de cap, gustul vinului sau senzația de roșeață a cerului de seară. (nota trad.)

entitate în viitor (un robot sau un avatar) care e complet convingătoare în reacțiile sale emoționale. Râde convingător la glumele tale și la rându-i te face să râzi și să plângi (însă nu doar ciupindu-te). Te convinge de sinceritatea sa când îți vorbește despre temerile și dorințele sale. Din toate punctele de vedere, entitatea pare conștientă. De fapt, pare să fie o persoană. Vei accepta că este o persoană conștientă?

Dacă reacția ta inițială este aceea că mai curând ai detecta vreun fel în care să-și trădeze natura nonbiologică, atunci nu respecți ipotezele acestei situații ipotetice, care au stabilit că este pe deplin convingătoare. Dată fiind această ipoteză, dacă ar fi amenințată de distrugere și ar răspunde așa cum ar face-o un om, arătându-se terorizată, ai reacționa în același mod empatic pe care l-ai manifesta dacă ai fi spectatorul unei scene în care ar fi implicat un om? În cazul meu, răspunsul este da și cred că răspunsul ar fi același pentru cei mai mulți, dacă nu pentru toți ceilalți oameni, indiferent de ce ar susține acum în cadrul unei dezbateri filosofice. Din nou, accentul, aici, cade pe cuvântul „convingătoare".

În mod cert, există dezacorduri privitoare la când sau dacă vom întâlni vreodată o asemenea entitate nonbiologică. Predicția mea consecventă este că acest lucru se va petrece pentru prima oară în anul 2029 și va deveni un fapt obișnuit în anii 2030. Dar, lăsând deoparte cadrul timpului, cred că vom ajunge în cele din urmă să considerăm că asemenea entități sunt conștiente. Să ne gândim cum le tratăm deja când avem de-a face cu ele în calitate de personaje ale poveștilor și filmelor: R2D2 din seria *Războiul Stelelor*, David și Teddy din filmul *A.I.*, Data din serialul de televiziune *Star Trek: The Next Generation*, Johnny 5 din filmul *Circuit*, WALL-E din filmul Disney *Wall-E*, T-800 – Terminatorul (cel bun) – din al doilea și ultimul film *Terminator*, Rachael Replica din filmul *Blade Runner* (care, apropo, nu este conștientă că nu e om), Bumblebee din filmul TV și seria comică *Transformers* și Sonny din filmul *I, Robot*. Empatizăm cu aceste personaje deși știm că sunt nonbiologice. Le considerăm ca fiind persoane conștiente, la fel cum facem cu personajele umane biologice. Le împărtășim sentimentele și ne temem pentru ele când intră în buclucuri. Dacă astăzi tratăm așa personajele nonbiologice de ficțiune, atunci tot așa vom trata în viitor și inteligențele din viața reală care se va întâmpla să nu aibă un substrat biologic.

Dacă accepți saltul de credință că o entitate nonbiologică pe care o găsești convingătoare în reacțiile sale față de experiențele subiective, conștiente, este într-adevăr conștientă, gândește-te doar la ce presupune asta:

respectiv, că această conștiență este o proprietate emergentă a modelului de ansamblu al entității, nu substratul pe baza căruia funcționează.

Există o prăpastie conceptuală între știință, care susține măsurarea obiectivă și concluziile pe care le tragem în acest fel, și conștiință, care este un sinonim pentru experiența subiectivă. În mod evident, nu putem să-i adresăm unei entități puse la îndoială întrebarea „Ești conștientă?" Dacă ne-am uita în „capul" ei – biologic sau ba – ca să ne asigurăm de răspuns, am avea de făcut presupuneri filosofice pentru a determina ce anume căutăm. Întrebarea dacă o entitate e sau nu conștientă nu este, prin urmare, una științifică. Pe această bază, unii observatori merg mai departe și pun sub semnul îndoielii faptul că conștiința însăși are vreo bază în realitate. Scriitorul și filosoful englez Susan Blackmore (născută în anul 1951) vorbește despre „marea iluzie a conștienței". Ea recunoaște realitatea memei (ideea) conștienței – cu alte cuvinte, conștiența există într-adevăr ca idee și există foarte multe structuri neocorticale care se ocupă de idee, ca să nu mai vorbim de cuvintele care au fost rostite și scrise despre ea. Dar nu este limpede dacă ea se referă la ceva real. Blackmore merge mai departe și explică faptul că ea nu neagă neapărat realitatea conștienței, ci mai curând încearcă să articuleze genurile de dileme cu care ne confruntăm atunci când încercăm să precizăm conceptul. După cum a scris psihologul și autorul britanic Stuart Sutherland (1927-1998) în *International Dictionary of Psychology – Dicționarul internațional de psihologie,* „conștiența este un fenomen fascinant, dar evaziv; este imposibil de specificat ce anume este, ce face sau de ce a evoluat".[4]

În orice caz, ar fi bine să nu respingem conceptul cu prea mare ușurință, considerându-l doar o dezbatere politicoasă între filosofi – dezbatere care, întâmplător, datează de două mii de ani, de la dialogurile platoniene. Ideea conștienței se află în sistemul nostru moral, iar sistemul nostru juridic, la rândul lui, este construit în mare pe aceste credințe morale. Dacă cineva stinge conștiența cuiva, cum se întâmplă în cazul unei crime, considerăm că acest act este imoral și, cu unele excepții, chiar o mare crimă. Aceste excepții sunt relevante și pentru conștiență, sub aspectul că putem autoriza forțele polițienești sau militare să omoare anumiți oameni conștienți pentru protejarea unui număr mai mare de alți oameni conștienți. Putem dezbate meritele excepțiilor, dar principiul de bază rămâne valabil.

Se consideră de asemenea că atacarea unui om și obligarea lui să îndure experiența suferinței este ceva imoral și ilegal. A-mi distruge

proprietatea este probabil un gest acceptabil. Dacă îți distrug ție proprietatea fără permisiunea ta, probabil că gestul nu este acceptabil, pentru că provoc suferință nu proprietății tale, ci ție, în calitatea ta de deținător al proprietății. Pe de altă parte, dacă proprietatea mea include o ființă conștientă, ca de pildă un animal, atunci eu, în calitate de deținător al acelui animal, nu am neapărat libertatea morală sau legală să fac cu el orice vreau – există, de exemplu, legi împotriva cruzimii față de animale.

Deoarece mare parte din sistemul nostru moral şi juridic se bazează pe protejarea existenței şi prevenirea suferințelor inutile provocate entităților conştiente, pentru a face judecăți responsabile trebuie să răspundem la întrebarea „Cine este conştient?" Prin urmare, această întrebare nu este doar o chestiune de dezbatere intelectuală, fapt evident în cazul controversei care învăluie un subiect cum ar fi avortul. Ar trebui să subliniez că subiectul avortului poate să depăşească pe undeva tema conştienței, partizanii proviață susținând că potențialul ca embrionul să devină în cele din urmă o persoană conştientă este un motiv suficient pentru a i se acorda acestuia protecție, la fel cum şi o persoană aflată în comă are acest drept. Dar, la nivelul fundamental, subiectul este o dezbatere cu privire la momentul în care fătul devine conştient.

Percepția conştienței ne afectează adeseori şi judecata în domenii controversate. Revenind la subiectul avortului, mulți oameni fac distincție între o măsură cum ar fi pilula din dimineața de după, care împiedică implantarea embrionului în uter în primele zile ale sarcinii, şi avortul în stadiul ulterior. Distincția se referă la posibilitatea ca fătul în stadiul ulterior să aibă conştiență. Este dificil să susții că un embrion cu vârsta de câteva zile este conştient dacă nu adopți o poziție panprotopsihistă, dar, chiar şi aşa, embrionul s-ar plasa sub cel mai simplu animal în termeni de conştiință. În mod similar, avem reacții foarte diferite la maltratarea maimuțelor mari, comparativ cu uciderea insectelor, de exemplu. Nimeni nu-şi face prea multe griji astăzi că ar provoca dureri şi suferințe software-ului computerului (deşi comentăm amănunțit capacitatea software-ului de a ne provoca nouă suferințe), dar, când software-ul viitorului va avea inteligența intelectuală, emoțională şi morală a omului biologic, această preocupare va deveni reală.

Prin urmare, poziția mea este că voi accepta entitățile nonbiologice care sunt pe deplin convingătoare în reacțiile lor emoționale drept persoane conştiente, iar predicția mea este că societatea le va accepta de asemenea. Observă, cititorule, că această definiție merge dincolo de entitățile care pot

trece Testul Turing, ce impune stăpânirea limbajului uman. Acestea din urmă sunt suficient de asemănătoare omului ca să le pot include şi cred că şi cea mai mare parte a societății le va include, însă eu adaug aici şi entitățile care manifestă reacții emoționale umane, dar pot să nu fie în măsură să treacă Testul Turing – de exemplu, copiii mici.

Răspunde asta la întrebarea filosofică cine este conştient, cel puțin pentru mine şi pentru alții care acceptă acest salt al credinței? Răspunsul este: nu chiar. Am acoperit doar un caz, cel al entităților care acționează de o manieră umană. Cu toate că discutăm despre viitoare entități, discutăm şi despre entități care demonstrează reacții umane convingătoare, deci această poziție este în continuare centrată pe om. Dar cum rămâne cu formele de inteligență mai deosebite care nu sunt asemănătoare omului? Ne putem imagina inteligențe la fel de complexe sau poate mult mai complexe şi mai complicate decât creierul omului, dar care au emoții şi motivații complet diferite. Cum să decidem dacă sunt sau nu conştiente?

Putem începe analizând ființe din lumea biologică al căror creier este comparabil cu al omului, dar, cu toate acestea, manifestă genuri de comportamente cu totul diferite. Filosoful britanic David Cockburn (născut în anul 1949) scrie despre urmărirea unei filmări cu un calmar gigant atacat (sau care credea că este atacat – Cockburn a emis ipoteza că animalul s-ar fi putut teme de omul cu camera de filmat). Calmarul s-a cutremurat şi s-a ferit, iar Cockburn scrie: „A reacționat într-un mod care m-a impresionat imediat şi puternic, deoarece era o manifestare a fricii. Partea şocantă în această succesiune a fost felul în care a fost posibil să văd în comportamentul unei ființe atât de diferite fizic de om o emoție care era, fără niciun fel de ambiguitate şi foarte precis, una de frică".[5] Autorul concluzionează că animalul a trăit această emoție şi îşi exprimă credința că oamenii care au văzut acel film au ajuns la aceeaşi concluzie în majoritatea lor. Dacă acceptăm descrierea şi concluzia lui Cockburn, atunci ar trebui să adăugăm pe lista noastră de entități conştiente calmarul gigant. Pe de altă parte însă, nici aşa nu am ajuns prea departe, deoarece concluzia se bazează pe reacția noastră empatică la o emoție pe care o recunoaştem în noi înşine. Este o perspectivă egocentrică sau umanocentrică.

Dacă ieşim din domeniul biologiei, inteligența nonbiologică va fi chiar şi mai variată decât inteligența din lumea biologică. De exemplu, e posibil ca unele entități să nu se teamă de propria distrugere şi să nu resimtă nevoia de emoții pe care o vedem la om sau la orice ființă biologică. Poate că

asemenea entități vor reuși să treacă Testul Turing sau poate că nici nu vor fi dornice să încerce.

De fapt, astăzi construim roboți lipsiți de simțul autoconservării pentru a realiza misiuni în medii periculoase. Încă nu sunt suficient de inteligenți sau de complecși pentru ca noi să fim obligați să ținem seama de sensibilitatea lor, dar ne putem imagina viitori roboți de acest gen la fel de complecși ca omul. Ce vom face în cazul lor?

Personal, aș spune că, dacă aș vedea în comportamentul unui asemenea dispozitiv devotamentul față de un scop complex și meritoriu și abilitatea de a lua decizii și a executa acțiuni notabile pentru îndeplinirea misiunii, aș fi impresionat și probabil că m-aș simți deranjat dacă el ar fi distrus. Poate că forțez puțin conceptul, în sensul că răspund la un comportament care nu include multe emoții considerate universale la om și chiar la ființele biologice de toate felurile. Însă, din nou, caut să mă conectez la atribute cu care pot relaționa în mine și în alți oameni. Ideea unei entități total devotate unui scop nobil și îndeplinirii lui sau cel puțin încercării de a face acest lucru fără să se gândească la binele său nu-i este, la urma urmelor, deloc străină experienței umane. În această clipă analizăm, de asemenea, o entitate care caută să protejeze oamenii biologici sau să ne promoveze cumva interesele.

Și dacă această entitate își are propriile scopuri, distincte de ale omului, și nu îndeplinește o misiune pe care am recunoaște-o ca fiind nobilă în termenii ce ne sunt proprii? Aș putea atunci să încerc să văd dacă aș putea lega de ceva sau aprecia unele dintre abilitățile ei într-un alt mod. Dacă într-adevăr este deosebit de inteligentă, e posibil să fie bună la matematică, așa că probabil aș putea să port cu ea o conversație pe acest subiect. Poate că mi-ar aprecia glumele matematice.

Dar dacă entitatea nu manifestă niciun interes să comunice cu mine, iar eu nu am suficient acces la acțiunile sale și la luarea deciziilor pentru a fi impresionat de frumusețea proceselor sale interne, conduc oare aceste lucruri la concluzia că entitatea nu este conștientă? Trebuie să concluzionez că entitățile care nu reușesc să mă convingă de reacțiile lor emoționale sau care nu sunt preocupate nici măcar să încerce s-o facă nu sunt în mod necesar lipsite de conștiență. Ar fi dificil să recunosc o altă entitate conștientă fără să stabilesc un anumit nivel de comunicare empatică, dar această judecată reflectă propriile mele limitări mai mult decât limitările entității în discuție. Prin urmare, trebuie să procedăm cu umilință. Este suficient de provocator să ne

punem, subiectiv, în locul altui om, aşa încât sarcina e cu atât mai dificilă în cazul inteligențelor care sunt extrem de diferite de a noastră.

De ce anume suntem conştienți?

> Dacă am putea privi prin craniu în creierul unei persoane care gândeşte în mod conştient şi dacă locul excitației optime ar fi luminos, atunci ar trebui să vedem jucând pe suprafața cerebrală o pată strălucitoare cu marginile fantastice, unduitoare, fluctuând permanent ca mărime şi formă, înconjurată de un întuneric mai mult sau mai puțin profund ce acoperă restul emisferei.
>
> **— Ivan Petrovici Pavlov, 1913**[6]

Revenind la calmarul gigant, putem recunoaşte unele dintre emoțiile sale aparente, dar cea mai mare parte din comportamentul său este un mister. Cum e să fii un calmar gigant? Cum se simte acesta când îşi strecoară corpul lipsit de coloană vertebrală printr-o deschizătură mică? Nici măcar nu avem vocabularul necesar pentru a răspunde la această întrebare, dat fiind faptul că nu putem descrie nici măcar experiențele pe care le împărtăşim cu alți oameni, de exemplu culoarea roşie sau senzația apei care ne stropeşte corpurile.

Dar nu trebuie să ajungem tocmai pe fundul oceanului pentru a găsi mistere în natura experiențelor conştiente – este suficient să ne gândim la ale noastre. Eu ştiu, de pildă, că sunt conştient. Eu presupun că tu, cititorule, eşti de asemenea conştient. (În ce-i priveşte pe oamenii care nu mi-au cumpărat cartea, nu sunt chiar atât de sigur.) Dar de ce anume sunt eu conştient? Ai putea să-ți pui şi tu, cititorule, această întrebare. Încearcă următorul experiment mintal (care va da rezultate în cazul acelora dintre voi care conduceți automobilul): imaginează-ți că te deplasezi pe banda stângă a autostrăzii. Acum închide ochii, pune mâinile pe volanul imaginar şi fă mişcările necesare pentru schimbarea benzii la dreapta.

În regulă, înainte să continui să citeşti, încearcă.

Iată ce-ai făcut, probabil: ai apucat volanul. Ai verificat dacă banda din dreapta este liberă. Presupunând că banda era liberă, ai rotit volanul la dreapta pentru scurt timp. Pe urmă, l-ai îndreptat la loc. S-a făcut.

E bine că nu ai fost într-o maşină adevărată, pentru că tocmai ai trecut peste toate benzile autostrăzii şi te-ai izbit de un copac. Poate că eu ar fi trebuit să-ţi spun să nu încerci asta într-un automobil real aflat în mişcare (dar presupun că, totuşi, ai înţeles deja regula că nu se conduce cu ochii închişi), deşi nu aceasta este aici problema-cheie. Dacă ai folosit procedura pe care tocmai am descris-o – şi aproape toată lumea o face în cazul acestui experiment mintal –, ai priceput totul greşit. Dacă mişti volanul la dreapta şi apoi îl aduci la poziţia iniţială, faci maşina să se îndrepte într-o direcţie diagonală pe direcţia ei de la început. Traversezi benzile din dreapta exact aşa cum ai intenţionat, dar continui să mergi la infinit spre dreapta până când ieşi de pe drum. Ceea ce trebuia să faci în timp ce maşina schimba banda spre dreapta era să roteşti volanul la stânga tot atât cât l-ai rotit la dreapta, iar apoi să-l aduci la poziţia iniţială. Asta ar fi făcut ca maşina să meargă drept înainte pe noua bandă.

Gândeşte-te că, dacă eşti şofer profesionist, ai făcut aceste mişcări de mii de ori. Nu eşti conştient când faci acest lucru? N-ai fost niciodată atent la ce faci atunci când schimbi benzile în mod real? Presupunând că nu citeşti această carte în spital, în convalescenţă după un accident produs la schimbarea benzilor, în mod cert stăpâneşti această abilitate. Dar, cu toate astea, nu eşti conştient de ce-ai făcut, chiar dacă ai făcut această mişcare de nenumărate ori.

Când îşi povestesc experienţele, oamenii le descriu ca succesiuni de situaţii şi decizii. Dar nu aşa trăim experienţa unei povestiri la început. Experienţa noastră iniţială reprezintă o succesiune de modele de nivel superior, dintre care o parte e posibil să fi generat senzaţii. În cel mai bun caz, ne amintim doar o mică parte din aceste modele. Chiar dacă suntem rezonabil de precişi în relatarea unei poveşti, ne folosim puterea de fabulaţie pentru a completa detaliile-lipsă şi a transforma succesiunea într-o naraţiune coerentă. Nu putem şti cu siguranţă care a fost experienţa noastră conştientă din amintirea ei; memoria este singura poartă de acces care ne poate aduce din nou faţă în faţă cu acea experienţă. Momentul actual este, să zicem, flotant şi se transformă rapid într-o amintire sau, cel mai adesea, nici atât. Chiar dacă o experienţă e transformată într-o amintire, aceasta este stocată, aşa cum indică Teoria minţii bazate pe recunoaşterea modelelor, ca model de nivel superior compus din alte modele, într-o ierarhie colosală. După cum am subliniat de mai multe ori, aproape toate experienţele pe care le trăim (cum ar fi ocaziile în care am schimbat benzile) sunt imediat uitate. Aşadar,

certitudinea în privința a ceea ce constituie experiența noastră conștientă este practic ceva de neatins.

Estul este est și vestul este vest

> Înainte să existe creierul, în Univers nu existau nici culori și nici sunete, nici mirosuri și nici arome și probabil că nu ar fi existat nici prea multe simțuri fără senzații sau emoții.
>
> **— Roger W. Sperry**[7]

> René Descartes intră într-un restaurant și se așază să ia cina. Ospătarul vine la el și îl întreabă dacă ar dori un aperitiv.
> – Nu, mulțumesc, spune Descartes. Aș vrea doar să comand cina.
> – Doriți să știți care sunt specialitățile zilei? întreabă ospătarul.
> – Nu, răspunde Descartes, care devine nerăbdător.
> – Doriți să beți ceva înainte de cină? Întreabă ospătarul.
> Descartes se simte insultat, deoarece e abstinent.
> – Nu prea cred! spune el indignat și, poof!, dispare.
>
> **— O glumă, așa cum mi-o amintesc, spusă de David Chalmers**

Există două moduri de a privi chestiunile pe care le-am luat în considerare – perspectivele corespondente, occidentale și răsăritene, asupra naturii conștienței și realității. În perspectiva apuseană, pornim de la o lume fizică în care au fost dezvoltate modele de informație. După câteva miliarde de ani de evoluție, entitățile din acea lume au evoluat suficient pentru a deveni ființe conștiente. În perspectiva răsăriteană, conștiența este realitatea fundamentală; lumea fizică prinde viață doar prin intermediul gândirii ființelor conștiente. Cu alte cuvinte, lumea fizică reprezintă gândurile devenite manifeste ale ființelor conștiente. Acestea sunt, desigur, doar simplificări ale filosofiilor complexe și diverse, dar întruchipează principalele polarități ale filosofiilor conștienței și ale relațiilor acesteia cu lumea fizică. Disensiunea Răsărit-Occident pe tema conștienței și-a mai găsit expresia și în școlile antagoniste de gândire din domeniul fizicii subatomice. În mecanica cuantică, particulele există într-o formă denumită câmp de probabilitate. Orice măsurătoare la care sunt supuse cu ajutorul unor dispozitive construite în acest scop

produce ceea ce se numeşte colapsul funcţiei de undă, asta însemnând că particula capătă, brusc, o anumită localizare. O concepţie larg răspândită precizează faptul că o asemenea măsurătoare implică observarea de către un observator conştient, deoarece, altminteri, măsurătoarea ar fi un concept lipsit de sens. Prin urmare, particula capătă o anumită localizare (ca şi alte proprietăţi, cum ar fi viteza) numai atunci când este observată. În esenţă, particulele se gândesc că, dacă tot nu e nimeni care să se uite la ele, nici ele nu trebuie să decidă unde se află. Eu numesc asta şcoala budistă de mecanică cuantică, deoarece, potrivit acesteia, particulele nu există înainte de a fi observate de o persoană conştientă.

Mai există o interpretare a mecanicii cuantice care evită o asemenea terminologie antropomorfică. În această analiză, câmpul ce reprezintă particula nu este unul de probabilitate, ci mai curând o funcţie care are valori diferite în locuri diferite. Prin urmare, câmpul se află, la modul fundamental, acolo unde se află şi particula. Există constrângeri la care valorile câmpului sunt supuse în diferite locuri, deoarece întregul câmp ce reprezintă o particulă acoperă doar un volum limitat de informaţie. De aici vine şi cuvântul „cuantic". Aşa-numitul colaps al funcţiei de undă, susţine această interpretare, nu este nici pe departe un colaps. Funcţia de undă nu dispare practic niciodată. Doar că instrumentul de măsură este făcut şi el din particule cu câmpuri, iar interacţiunea câmpului particulei care e măsurată cu câmpurile particulelor instrumentului de măsură se concretizează prin citirea faptului că particula s-ar afla într-un anumit loc. Însă câmpul este prezent în continuare. Aceasta e interpretarea apuseană a mecanicii cuantice, deşi merită observat faptul că punctul de vedere mai popular în rândurile fizicienilor din lumea întreagă îl reprezintă ceea ce am numit eu *interpretarea răsăriteană*.

A existat un filosof ale cărui scrieri au reparat această fractură Răsărit-Apus. Gânditorul britanic de origine austriacă Ludwig Wittgenstein (1889–1951) a studiat filosofia limbii şi a cunoaşterii şi a cugetat îndelung la întrebarea „Ce putem şti cu adevărat?" A cumpănit acest subiect ca soldat în Primul Război Mondial şi a făcut notiţe pentru ceea ce avea să fie singura sa carte publicată în timpul vieţii, *Tractatus Logico-Philosophicus.* Lucrarea avea o structură neobişnuită şi abia prin eforturile fostului său mentor, matematicianul şi filosoful britanic Bertrand Russell, şi-a găsit un editor în anul 1921. A devenit Biblia unei şcoli majore de filosofie, cunoscută sub denumirea de pozitivism logic, care urmărea să definească limitele ştiinţei. Cartea şi mişcarea care a înconjurat-o l-au influenţat pe Turing şi apariţia teoriei computaţiei şi

lingvisticii. *Tractatus Logico-Philosophicus* intuiește că întreaga cunoaștere este ierarhică. De fapt, cartea însăși este construită din afirmații înseriate și numerotate. De exemplu, primele patru afirmații ale cărții sunt:

1. Lumea este tot ceea ce e cazul.
1.1. Lumea este totalitatea faptelor, nu a lucrurilor.
1.11. Lumea este determinată de fapte și de faptul că toate sunt fapte.
1.12. Căci totalitatea faptelor determină ce e cazul și, de asemenea, ce nu e cazul.

O altă afirmație semnificativă din *Tractatus* – și cea căreia Turing i-a devenit ecoul – este aceasta:

4.0031. Întreaga filosofie este o critică a limbii.

În esență, atât *Tractatus Logico-Philosophicus*, cât și mișcarea pozitivismului logic afirmă că realitatea fizică există separat de perceperea realității fizice pe care o deținem noi, dar că tot ceea ce putem ști despre realitate e ceea ce percepem cu ajutorul simțurilor – care pot fi îmbogățite cu ajutorul instrumentelor noastre – și cu ajutorul deducțiilor logice pe care le putem realiza în urma acestor impresii senzoriale. În esență, Wittgenstein încearcă să descrie metodele și scopurile științei.

Afirmația finală din carte are numărul 7: „Ceea ce nu putem vorbi trebuie transmis mai departe în tăcere". În consecință, acel Wittgenstein de la început considera discuția despre conștiență ca fiind circulară și tautologică – prin urmare, o risipă de timp.

Însă celălalt Wittgenstein, de mai târziu, respinge complet această abordare și acordă întreaga sa atenție filosofică discuției despre chestiunile în legătură cu care argumentase anterior că trebuie să fie transmise mai departe în tăcere. Lucrările lui pe tema acestei gândiri revizuite au fost adunate și publicate în 1953, la doi ani după moartea filosofului, într-o carte intitulată *Philosophical Investigations – Investigații filosofice*. Și-a criticat ideile anterioare din *Tractatus*, considerându-le circulare și goale de conținut, și a ajuns la concepția că lucrurile asupra cărora recomandase să se păstreze tăcerea erau, de fapt, chiar cele asupra cărora merită să reflectăm. Aceste scrieri i-au influențat profund pe existențialiști, făcând din Wittgenstein unicul personaj

din filosofia modernă care a devenit arhitectul de prim rang al unui număr de două şcoli contradictorii de gândire filosofică.

Dar despre ce anume considera Wittgenstein cel de mai târziu că merită să se cumpănească şi să se vorbească? Despre subiecte precum frumuseţea şi iubirea, pe care le recunoştea ca existând sub formă imperfectă de idei în minţile oamenilor. În acelaşi timp însă, filosoful scrie că asemenea concepte există cu adevărat într-o sferă perfectă şi idealizată, similară „formelor" perfecte despre care a scris Platon în *Dialoguri*, o altă lucrare care a adus lumină asupra abordărilor naturii realităţii, aparent contradictorie.

Un gânditor a cărui poziţie cred că e greşit etichetată este filosoful şi matematicianul francez René Descartes. Celebra sa formulă „Gândesc, deci exist" este interpretată la modul general ca o preamărire a gândirii raţionale, în sensul că „Eu gândesc, adică pot să exercit gândirea raţională, prin urmare sunt valoros". Pe cale de consecinţă, Descartes este considerat arhitectul perspectivei raţionale apusene.

Pe de altă parte însă, citind această afirmaţie în contextul altor scrieri, am o altă impresie. Descartes a fost tulburat de ceea ce se numeşte „problema minte-corp". Mai exact, cum este posibil să apară o minte conştientă din materia fizică a creierului? Potrivit acestei perspective, s-ar părea că încerca să împingă scepticismul raţional la punctul de fractură, aşa că, din punctul meu de vedere, afirmaţia lui înseamnă, de fapt, „Gândesc, respectiv se petrece o experienţă subiectivă, prin urmare tot ce ştiu cu siguranţă e că există ceva – să-i spunem Eu". Nu a putut fi sigur că lumea fizică există, pentru că tot ce avem la dispoziţie sunt impresiile simţurilor individuale determinate de ea, care pot fi greşite sau complet iluzorii. Pe de altă parte, însă, ştim că există experimentatorul.

Am fost crescut în sânul Bisericii Unitariene, în care am studiat toate religiile lumii. Studiam, să spunem, şase luni religia budistă, apoi treceam la serviciul religios budist, citeam cărţile religiei budiste şi purtam discuţii de grup cu liderii budişti. Pe urmă ne mutam la o altă religie, cum ar fi iudaismul. Tema generală era „numeroasele căi către adevăr", laolaltă cu toleranţa şi transcendenţa. Această din urmă idee însemna că rezolvarea aparentelor contradicţii dintre tradiţii nu impune luarea deciziei că una este corectă, iar cealaltă greşită. Adevărul poate fi descoperit numai găsindu-se o explicaţie care elimină – transcende – aparentele diferenţe, în special în cazul întrebărilor fundamentale legate de semnificaţie şi scop.

Așa rezolv eu diferendele Apus-Răsărit privitoare la conștiență și la lumea fizică. Din punctul meu de vedere, ambele perspective trebuie să fie adevărate.

Pe de o parte, este o prostie să se nege lumea fizică. Chiar dacă trăim într-o simulare, cum specula filosoful suedez Nick Bostrom, realitatea este, totuși, pentru noi un nivel conceptual real. Dacă acceptăm existența lumii fizice și evoluția care s-a petrecut în sânul ei, atunci putem vedea că din ea au evoluat entități conștiente.

Pe de altă parte, perspectiva răsăriteană – potrivit căreia conștiența este fundamentală și reprezintă unica realitate cu adevărat importantă – e de asemenea dificil de negat. Gândește-te numai, cititorule, la respectul deosebit pe care îl acordăm persoanelor conștiente față de cel acordat lucrurilor neconștiente. Considerăm că acestea din urmă nu au nicio valoare intrinsecă în afara măsurii în care pot influența experiența subiectivă a persoanelor conștiente. Chiar dacă înțelegem conștiența ca pe o proprietate emergentă a unui sistem complex, nu putem spune că ar fi vorba doar de un alt atribut (laolaltă cu „digestia" și „lactația", ca să-l citez pe John Searle). Reprezintă ceea ce e cu adevărat important.

Cuvântul „spiritual" este adeseori folosit pentru a denumi lucruri cu o semnificație supremă. Multor oameni nu le place să folosească o asemenea terminologie provenită din tradițiile spirituale sau religioase pentru că implică un set de credințe cu care s-ar putea să nu fie de acord. Dar, dacă eliminăm complexitatea mistică a tradițiilor religioase și respectăm pur și simplu termenul „spiritual" ca pe unul care implică un fapt cu semnificație profundă pentru om, conceptul de conștiență este foarte potrivit. El reflectă suprema valoare spirituală. Pe bună dreptate, însuși „spiritul" este adeseori folosit pentru a denumi conștiența.

Atunci, evoluția poate fi considerată un proces spiritual prin faptul că creează ființe spirituale, respectiv entități conștiente. Iar evoluția se îndreaptă către o și mai mare complexitate, o și mai mare cunoaștere, o și mai mare inteligență, o și mai mare frumusețe, o și mai mare creativitate și abilitate de a exprima emoții mai transcendentale, cum ar fi iubirea. Toate acestea sunt descrieri pe care oamenii le-au folosit pentru conceptul de Dumnezeu, cu toate că Dumnezeu este descris ca neavând limite din aceste puncte de vedere.

Oamenii se simt adeseori amenințați de discuții care presupun posibilitattea ca o mașină să fie conștientă, deoarece au impresia că orice

considerent în acest sens ar fi o denigrare a valorii spirituale a persoanei conştiente. Dar această reacţie reflectă o înţelegere greşită a conceptului de maşină. Criticii de acest gen abordează chestiunea pe baza maşinilor pe care le cunosc acum şi, oricât de impresionante ar deveni acestea, sunt de acord că exemplele tehnologice contemporane nu sunt încă demne de respectul nostru ca fiinţe conştiente. Predicţia mea este că vor deveni de nedistins de omul biologic, pe care îl considerăm într-adevăr o fiinţă conştientă şi, prin urmare, se vor împărtăşi din valoarea spirituală pe care o atribuim conştienţei. Nu este vorba despre o discreditare a omului; dimpotrivă, este o înălţare a felului în care înţelegem (unele dintre) maşinile viitoare. Probabil că ar trebui să adoptăm o terminologie diferită pentru aceste entităţi, de vreme ce vor fi maşini de un alt gen.

Pe bună dreptate, când privim acum în interiorul creierului şi îi decodificăm mecanismele, descoperim metode şi algoritmi pe care nu-i putem doar înţelege, ci îi putem şi recrea – „părţi ale unei prese care se împing unele pe altele", ca să-l parafrazez pe matematicianul şi filosoful german Gottfried Wilhelm Leibniz (1646-1716) când a scris despre creier. Omul alcătuieşte deja maşini spirituale. Mai mult de-atât, ne vom contopi cu instrumentele pe care le creăm într-atât, încât distincţia dintre om şi maşină se va estompa până când va dispărea. Acest proces este deja în curs de desfăşurare, chiar dacă majoritatea maşinilor care ne devin prelungiri încă nu se află în interiorul trupurilor şi creierelor noastre.

Liberul arbitru

> Un aspect central al conştienţei este capacitatea de a privi înainte, pe care o numim „anticipare". Este capacitatea de a planifica şi, în termeni sociali, de a schiţa un scenariu pentru ceea ce este posibil să se întâmple sau se va întâmpla în mod cert în interacţiunile sociale care nu s-au produs încă. Prin acest sistem ne îmbunătăţim şansele de a face acele lucruri care vor reprezenta interesul nostru... Sugerez că „liberul arbitru" reprezintă aparenta noastră capacitate de a alege şi acţiona aşa cum ni se pare cel mai util sau apropriat şi, totodată, insistenţa noastră că aceste alegeri ne aparţin.
>
> **— Richard D. Alexander**

Vom spune că planta nu ştie ce se petrece pentru simplul motiv că nu are ochi, urechi sau creier? Dacă spunem că acționează mecanic sau numai mecanic, nu vom fi forțați să admitem că şi alte felurite acțiuni aparent deliberate sunt de asemenea mecanice? Dacă nouă ni se pare că planta ucide şi mănâncă o muscă în mod deliberat, plantei nu ar trebui să i se pară că omul trebuie să ucidă şi să mănânce o oaie în mod mecanic?

— Samuel Butler, 1871

Este creierul, care e remarcabil de dublu ca structură, un organ dublu, „aparent separat, dar totuşi o unitate în separație"?

— Henry Maudsley[8]

Redundanța, după cum am aflat, este o strategie-cheie folosită de neocortex. Dar în creier mai există un nivel de redundanță, în sensul că emisferele sale stângă şi dreaptă, deşi nu sunt identice, sunt în mare măsură similare. La fel cum anumite regiuni ale neocortexului ajung în mod normal să proceseze anumite tipuri de informații, emisferele se specializează şi ele – de exemplu, emisfera stângă e răspunzătoare de obicei de limbajul verbal. Dar aceste alocări pot să fie şi rerutate, până în punctul în care putem supraviețui şi funcționa oarecum normal cu o singură jumătate de creier. Stella de Bode şi Susan Curtiss, cercetători americani în domeniul neuropsihologiei, au comunicat un număr de patruzeci şi nouă de copii care au fost supuşi unei hemisferotomii (extirparea unei jumătăți din creier), intervenție chirurgicală extremă, aplicată pacienților care suferă de atacuri cerebrale ce le pun viața în pericol şi se manifestă la nivelul unei singure emisfere. Unii copii supuşi acestei proceduri rămân cu deficiențe, dar deficiențele sunt specifice, iar pacienții au personalități rezonabil de normale. Mulți dintre ei se dezvoltă, aşa încât celor care îi văd nu le este deloc evident că au doar o jumătate de creier. De Bode şi Curtiss scriu despre copii hemisferotomizați pe stânga care „îşi dezvoltă un limbaj remarcabil de bun, în ciuda extirpării «emisferei limbajului»".[9] Autoarele descriu un asemenea student care a terminat colegiul, a urmat cursuri postuniversitare şi a obținut scoruri peste medie la testele IQ. Studiile efectuate au indicat efecte minime pe termen lung asupra cogniției de ansamblu, memoriei, personalității şi simțului umorului.[10] În cursul unui studiu realizat în anul 2007, cercetătorii americani Shearwood McClelland şi Robert Maxwell au adus la lumină rezultate la fel de pozitive pe termen lung la adulți.[11]

S-a comunicat, de asemenea, că o fetiță de zece ani din Germania, care s-a născut cu o jumătate de creier, era cât se poate de normală. Avea vederea aproape perfectă pentru un ochi, în vreme ce pacienții care sunt supuși unei intervenții chirurgicale de hemisferotomie își pierd o parte din câmpul vizual imediat după operație.[12] Cercetătorul scoțian Lars Muckli a comentat: „Creierul are o plasticitate uluitoare, dar am fost cât se poate de uluiți când am văzut cât de bine s-a adaptat emisfera stângă a creierului acestei fetițe pentru compensarea jumătății-lipsă".

Deși aceste observații sprijină cu siguranță ideea plasticității neocortexului, implicația lor mai interesantă este aceea că fiecare părem a avea două creiere, nu unul singur, și că ne putem descurca foarte bine cu oricare dintre ele. Dacă îl pierdem pe unul, pierdem modelele corticale care sunt stocate exclusiv în el, dar fiecare creier este, în sine, cât se poate de complet. Prin urmare, fiecare emisferă își are o conștiență proprie? Se poate argumenta că așa stau lucrurile.

Să ne gândim la pacienții cu scindare cerebrală, care au în continuare cele două emisfere ale creierului, dar canalul dintre ele este secționat. Corpul calos e un mănunchi format din aproximativ 250 de milioane de axoni care interconectează emisferele cerebrale stângă și dreaptă și le permite acestora să comunice și să se coordoneze una cu cealaltă. La fel cum doi oameni pot să comunice între ei strâns și să acționeze ca un singur factor de decizie, deși rămân doi indivizi separați și completi, cele două emisfere ale creierului pot funcționa ca o unitate, rămânând totuși independente.

Așa cum sugerează și termenul, la pacienții care suferă de scindare cerebrală corpul calos a fost tăiat sau mutilat, aceștia rămânând efectiv cu două creiere funcționale, fără comunicare directă între ele. Michael Gazzaniga, cercetător american din domeniul psihologiei (născut în anul 1939), a realizat experimente extensive pentru a afla ce gândește fiecare emisferă cerebrală a unui pacient cu scindare cerebrală.

Emisfera stângă a unui asemenea pacient vede în general câmpul vizual drept și invers. Gazzaniga și colegii săi i-au arătat unui subiect cu scindare cerebrală imaginea unei gheare de pui în câmpul vizual drept (imagine care era văzută de emisfera lui stângă) și o scenă cețoasă în câmpul vizual stâng (care era văzută de emisfera dreaptă). Pe urmă i-au prezentat o colecție de fotografii, astfel încât acestea să poată fi văzute de ambele emisfere, și i-au cerut pacientului să aleagă una dintre fotografiile care se potriveau cu prima imagine. Mâna stângă a pacientului (controlată

de emisfera lui dreaptă) a indicat fotografia unei lopeți, în vreme ce mâna dreaptă a indicat imaginea unui pui. Până acum, toate bune – cele două emisfere acționau independent şi logic. „De ce ai ales asta?", l-a întrebat Gazzaniga pe pacient, care a răspuns verbal (controlat de centrul vorbirii din emisfera stângă): „Gheara de pui se potriveşte în mod evident cu puiul". Dar apoi pacientul a privit în jos şi, observând că mâna lui stângă arăta lopata, a explicat imediat acest fapt (din nou centrul vorbirii din emisfera stângă): „Şi ai nevoie de o lopată ca să cureți mizeria de la pui".

Este o născocire. Emisfera dreaptă (care controlează brațul şi mâna stângă) indică în mod corect lopata, dar, cum emisfera stângă (care controlează răspunsul verbal) nu este conştientă de zăpadă, născoceşte o explicație, fără a fi conştientă că povesteşte. Îşi asumă responsabilitatea pentru o acțiune pe care nu a decis-o şi nu a întreprins-o niciodată, dar crede că a făcut-o.

Asta implică faptul că fiecare dintre cele două emisfere ale unui pacient cu scindare cerebrală îşi are propria conştiență. Emisferele par a nu fi conştiente că trupul lor este controlat, efectiv, de două creiere, deoarece acestea învață să se coordoneze între ele, iar deciziile lor sunt suficient de bine aliniate şi consecvente pentru ca fiecare să creadă că deciziile celeilalte îi aparțin.

Experimentul lui Gazzaniga nu dovedeşte că un individ normal, cu un corp calos funcțional, are două jumătăți de creier conştiente, dar este sugestiv în privința acestei posibilități. Corpul calos permite o colaborare eficientă între cele două jumătăți de creier, dar asta nu înseamnă neapărat că nu sunt nişte minți separate. Fiecare poate fi păcălită şi făcută să creadă că a luat toate deciziile, deoarece acestea ar fi suficient de apropiate față de ce ar decide fiecare în mod independent şi, la urma urmelor, fiecare are suficient de multă influență asupra oricărei decizii (grație coroborării cu cealaltă emisferă prin intermediul corpului calos). Astfel, fiecăreia dintre cele două minți i s-ar părea că deține controlul.

Cum ai testa situația în care ambele sunt conştiente? Ele pot fi evaluate sub aspectul corelării neurologice a conştienței, adică exact ce a făcut Gazzaniga. Experimentele acestuia arată că fiecare emisferă acționează ca un creier separat. Povestirea nu este limitată la emisferele creierului; fiecare dintre noi facem aşa ceva în mod curent. Fiecare emisferă este aproape la fel de inteligentă ca omul, deci, dacă credem că creierul omului este conştient, atunci trebuie să concluzionăm că fiecare emisferă este conştientă

în mod independent. Putem evalua corelările neurologice şi putem realiza propriile noastre experimente mintale (de exemplu, considerând că, dacă două emisfere cerebrale fără un corp calos funcțional constituie două minți conştiente separate, acelaşi lucru e valabil şi pentru două emisfere cu conexiune funcțională între ele), dar orice încercare de detectare mai directă a conştienței fiecărei emisfere ne pune din nou față în față cu lipsa unui test ştiințific pentru conştiență. Iar dacă acceptăm că fiecare emisferă a creierului este conştientă, atunci încuviințăm şi că aşa-numita activitate inconştientă de la nivelul neocortexului (care constituie cea mai mare parte a activității sale) are de asemenea o conştiență independentă? Sau poate că are mai multe, nu doar una. Pe bună dreptate, Marvin Minsky se referă la creier ca la o „societate a minții".[13]

În cadrul unui alt experiment pe un creier scindat, cercetătorii au arătat cuvântul „clopot" creierului drept şi cuvântul „muzică" creierului stâng. Pacientul a fost întrebat ce cuvânt vede. Centrul vorbirii controlat de emisfera stângă a spus „muzică". Subiectului i-a fost prezentat apoi un grup de fotografii şi i s-a cerut să indice imaginea cel mai strâns legată de cuvântul care abia îi fusese arătat. Brațul controlat din emisfera dreaptă a arătat clopotul. Când a fost întrebat de ce a indicat clopotul, centrul vorbirii care controlează vorbirea, situat în emisfera stângă, a replicat: „Ei bine, muzica, ultima oară când am auzit o muzică a fost când au bătut clopotele aici, afară". A dat această explicație deşi mai existau şi alte fotografii dintre care putea să aleagă, mult mai strâns legate de muzică.

Din nou, este vorba de o născocire. Emisfera stângă explică de parcă ar fi decizia ei, pe care însă nu a luat-o niciodată şi nici nu a pus-o vreodată în aplicare. Nu face acest lucru pentru a-şi acoperi un prieten (respectiv, cealaltă emisferă), ci crede realmente că decizia i-a aparținut.

Aceste reacții şi decizii se pot extinde şi la răspunsurile emoționale. Un pacient adolescent cu creierul scindat – astfel încât ambele emisfere auzeau – a fost întrebat „Cine este ... preferată" şi pe urmă i s-a servit cuvântul „iubita" doar pentru emisfera dreaptă, prin urechea stângă. Gazzaniga relatează că subiectul a roşit şi a avut un comportament jenat, o reacție adecvată pentru un adolescent care este întrebat despre iubita lui. Însă centrul vorbirii din emisfera stângă a relatat că nu auzise niciun cuvânt şi a cerut o clarificare: „Cine anume e preferată?" Când i s-a cerut din nou să răspundă la întrebare, de această dată în scris, mâna stângă, controlată de emisfera dreaptă, a scris numele iubitei sale.

Testele lui Gazzaniga nu sunt experimente mintale, ci veritabile experimente ale minții. Deși oferă o perspectivă interesantă asupra subconștientului, ele vorbesc și mai direct despre liberul arbitru. În fiecare dintre aceste cazuri, una dintre emisfere crede că a luat o decizie pe care, în realitate, n-a luat-o. În ce măsură este adevărat acest lucru pentru deciziile pe care le luăm în fiecare zi?

Să vedem cazul unei paciente, o fetiță de zece ani bolnavă de epilepsie. Neurochirurgul Itzhak Fried a realizat o intervenție chirurgicală cu pacienta trează (ceea ce este fezabil, deoarece la nivelul creierului nu există receptori ai durerii).[14] De câte ori el stimula un anumit punct de pe neocortex, ea începea să râdă. La început, echipa de chirurgi s-a gândit că era posibil să declanșeze un fel de reflex al râsului, dar și-a dat repede seama că stimula, practic, percepția umorului. Părea că descoperiseră un punct de pe neocortex – în mod evident există mai multe, nu doar unul – care recunoaște percepția umorului. Fetița nu doar că râdea – situația chiar i se părea comică, deși nu se schimbase nimic în stare ei în afara faptului că îi era stimulat acel punct al neocortexului. Când medicii au întrebat-o de ce râde, nu numai că a răspuns în sensul general „A, nu, fără vreun motiv anume" sau „Tocmai mi-ați stimulat creierul", ci a și născocit imediat un motiv. A arătat ceva din încăpere și a încercat să explice de ce era amuzant lucrul respectiv. „Sunteți atât de amuzanți, cum stați așa, acolo", a fost comentariul ei tipic.

Aparent, suntem foarte dornici să ne explicăm și să ne raționalizăm acțiunile, chiar și când nu noi suntem autorii deciziilor care au condus la ele. Așadar, cât de responsabili suntem de deciziile noastre? Să ne gândim la următoarele experimente ale lui Benjamin Libet (1916–2007), profesor de fiziologie la Universitatea California din Davis. Libet le-a cerut participanților să se așeze în fața unui cronometru, cu electrozi de electroencefalografie atașați pe cap. I-a instruit să efectueze sarcini simple, cum ar fi să apese un buton sau să miște un deget. Participanților li s-a cerut să noteze timpul de pe cronometru „prima oară când deveneau conștienți de dorința sau imboldul de a acționa". Testele au indicat o marjă de eroare de numai 50 de milisecunde la aceste autoevaluări realizate de subiecți. De asemenea, au cronometrat o medie de 200 de milisecunde între momentul în care subiecții și-au exprimat conștientizarea nevoii presante de a acționa și actul propriu-zis.[15]

Cercetătorii au urmărit de asemenea și semnalele EEG provenite de la creierele subiecților. Activitatea cerebrală implicată în inițierea acțiunii de către cortexul motor (care este răspunzător de realizarea acțiunii) s-a

manifestat practic, în medie, cu 500 de milisecunde înainte de îndeplinirea sarcinii. Asta înseamnă următorul lucru: cortexul motor se pregătea să îndeplinească sarcina cu aproximativ o treime de secundă înainte ca subiectul să fie conştient măcar că a luat decizia de a face lucrul respectiv.

Implicaţiile experimentelor lui Libet au fost dezbătute intens. Libet a concluzionat că însăşi conştientizarea de către noi a luării unei decizii pare să fie o iluzie şi că, până la urmă, „conştienţa nu intră în buclă". Filosoful Daniel Dennett a comentat: „Acţiunea este iniţial precipitată într-o anumită parte a creierului şi de acolo pornesc semnalele pentru muşchi, oprindu-se în drum să-ţi spună ţie, agentul conştient, ce se petrece (dar, ca toţi bunii oficiali, lăsându-te pe tine, preşedintele mult-prea-important, să rămâi cu iluzia că tu ai început tot)".[16] În acelaşi timp, Dennett a pus sub semnul îndoielii intervalele de timp înregistrate de experiment, în esenţă susţinând că era posibil ca subiecţii să nu fi fost, de fapt, conştienţi de momentul în care au devenit conştienţi de decizia de a acţiona. Te-ai putea întreba: dacă subiectul nu este conştient când devine conştient de luarea unei decizii, atunci cine e conştient? Însă aspectul fundamental este surprins foarte bine – după cum am arătat anterior, motivul pentru care suntem conştienţi e departe de a fi clar.

Vilayanur Subramanian „Rama" Ramachandran (născut în anul 1951), american de origine indiană, specialist în neuroştiinţă, explică situaţia puţin diferit. Dat fiind faptul că la nivelul neocortexului avem aproximativ 30 de miliarde de neuroni, aici se petrec permanent foarte multe lucruri, dar suntem conştienţi de foarte puţine dintre acestea. Deciziile, mari şi mici, sunt procesate continuu de neocortex, iar soluţiile propuse ies la iveală în conştienţa noastră. În locul liberului arbitru, Ramachandran sugerează că ar trebui să vorbim despre „libera opoziţie" – respectiv forţa de a respinge soluţiile propuse de părţile nonconştiente ale neocortexului.

Să ne gândim la analogia cu o campanie militară. Oficialii militari pregătesc o recomandare pentru preşedinte. Anterior primirii aprobării de la acesta, ei realizează o activitate pregătitoare care să permită punerea deciziei în aplicare. Într-un anumit moment, decizia propusă îi este prezentată preşedintelui, care o aprobă şi apoi se întreprinde restul misiunii. Deoarece „creierul" reprezentat de această analogie implică atât procesele inconştiente de la nivelul neocortexului (respectiv, oficialii de sub preşedinte), cât şi procesele sale conştiente (preşedintele), am vedea atât activitatea neuronală, cât şi acţiunile propriu-zise ca desfăşurându-se anterior luării

deciziei oficiale. Întotdeauna, într-o anumită situație, se poate ajunge la dezbateri legate de mărimea răgazului pe care oficialii de sub președinte i-l acordă acestuia pentru acceptarea sau respingerea unei recomandări și în mod cert președinții Americii au făcut și una, și alta. Însă nu ar trebui să ne surprindă faptul că activitatea mintală, chiar și la nivelul cortexului motor, poate începe înainte ca noi să fim conștienți că exista o decizie care trebuia luată.

Experimentele lui Libet subliniază că, la nivelul creierului nostru, există o activitate foarte intensă, care stă la baza deciziilor noastre și care nu este conștientă. Știam deja că mare parte din ceea ce se petrece în neocortex nu este conștient; prin urmare, nu ar trebui să fie surprinzător faptul că acțiunile și deciziile noastre își au originea atât în activitatea inconștientă, cât și în cea conștientă. Este importantă această distincție? Dacă deciziile noastre rezultă și din una, și din alta, ar trebui să conteze dacă separăm partea conștientă de cea inconștientă? Nu cumva ne reprezintă creierul ambele aspecte? Nu suntem, oare, în cele din urmă, responsabili pentru tot ceea ce se petrece în creierul nostru? „Da, am tras în victimă, dar nu sunt responsabil, pentru că nu eram atent" probabil că este o apărare slabă. Deși există unele temeiuri juridice înguste pe baza cărora o persoană nu poate fi făcută răspunzătoare pentru deciziile sale, în general suntem făcuți răspunzători pentru toate alegerile pe care le facem.

Observațiile și experimentele pe care le-am citat mai înainte constituie experimente mintale privitoare la liberul arbitru, un subiect care, asemenea celui al conștienței, a fost dezbătut încă de la Platon încoace. Însăși sintagma „liber arbitru" datează încă din secolul al treisprezecelea, dar ce anume înseamnă ea mai exact?

Dicționarul Merriam-Webster o definește ca fiind „libertatea omului de a face alegeri care nu sunt determinate de cauze anterioare sau de intervenția divină". Vei observa, cititorule, că această definiție este fără speranță de circulară: „Liberul arbitru este libertatea..." Lăsând deoparte ideea unei intervenții divine care ar sta împotriva liberului arbitru, există un element util al acestei definiții – și anume ideea că decizia „nu este determinată de cauze anterioare". Voi reveni imediat la asta.

Enciclopedia Stanford de filosofie afirmă că liberul arbitru este „capacitatea agenților raționali de a alege un curs al acțiunii dintre mai multe alternative". Potrivit acestei definiții, un simplu computer este capabil de liber arbitru, astfel încât e mai puțin util decât definiția de dicționar.

Wikipedia este, de fapt, ceva mai bună. Defineşte liberul arbitru ca fiind „abilitatea agenților de a face alegeri libere de anumite genuri de constrângeri... Constrângerea care a generat preocuparea dominantă a fost... determinismul". Din nou se foloseşte cuvântul „liber" pentru a defini liberul arbitru, dar nu se lămureşte ceea ce a fost considerat ca fiind principalul dușman al liberului arbitru: determinismul. Din acest punct de vedere, definiția din Merriam-Webster de mai înainte este practic similară referinței la deciziile care „nu sunt determinate de cauze anterioare".

Aşadar, ce vrem să spunem prin determinism? Dacă introduc „2 + 2" într-un calculator şi acesta afişează rezultatul „4", pot spune că acest calculator şi-a afişat liberul arbitru decizând să afişeze acel „4"? Nimeni nu ar accepta asta ca demonstrare a liberului arbitru, deoarece „decizia" a fost predeterminată de mecanismele interne ale calculatorului şi de input. Dacă introduc un calcul mai complicat, ajung la aceeaşi concluzie în privința lipsei de liber arbitru.

Dar Watson atunci când dă răspunsurile la întrebările de la *Jeopardy*!? Cu toate că deliberările lui sunt mult mai complexe decât cele ale calculatorului, foarte puțini observatori ar atribui deciziile sale liberului arbitru. Niciun om nu ştie cum funcționează programele lui, dar putem identifica un grup de oameni capabili să-i descrie toate metodele. Mai important, output-ul lui Watson este determinat de (1) toate programele lui, în momentul în care este formulată întrebarea; (2) întrebarea propriu-zisă; (3) starea parametrilor interni care îi influențează deciziile şi (4) miile de miliarde de bytes de baze de cunoştințe, inclusiv enciclopedii. Output-ul lui este determinat pe baza acestor patru categorii de informații. Am putea să speculăm că adresarea aceleiaşi întrebări ar aduce întotdeauna acelaşi răspuns, dar Watson e programat să învețe din experiențele sale, astfel încât există posibilitatea ca răspunsurile ulterioare să fie diferite. Pe de altă parte, asta nu contrazice analiza pe care o fac acum; dimpotrivă, constituie doar o schimbare la punctul 3, parametrul care îi controlează lui Watson deciziile.

Aşadar, prin ce se deosebeşte, mai exact, un om de Watson, astfel încât omului să-i putem atribui liber arbitru, pe când programului de computer nu? Cu toate că Watson este un jucător mai bun de *Jeopardy!* comparativ cu majoritatea oamenilor, nu e nici pe departe la fel de complex ca un neocortex uman. Watson posedă, într-adevăr, foarte multe cunoştințe şi foloseşte metode ierarhice, însă complexitatea gândirii lui ierarhice e în continuare considerabil mai redusă decât cea a omului. Deci diferența se referă

la scara de complexitate a gândirii sale ierarhice? Se poate argumenta că problema nu se reduce la asta. În discuția mea referitoare la subiectul conștienței, am menționat că propriul meu salt de gândire rezidă în faptul de a considera că un computer care a trecut Testul Turing este conștient. Nici cele mai bune *chatbot*-uri nu sunt capabile, astăzi, să facă acest lucru (deși sunt permanent îmbunătățite), astfel încât concluzia mea cu privire la conștiență este o chestiune de nivel de performanță al entității. Poate că același lucru este adevărat și în legătură cu faptul că îi atribui liber arbitru.

Conștiența reprezintă, într-adevăr, o diferență filosofică între creierul omului și programele software contemporane. Considerăm creierul omului ca fiind conștient, dar nu atribuim – încă – aceeași calitate și programelor software. Să fie acesta factorul pe care-l căutăm și care stă la baza liberului arbitru?

Un simplu experiment mintal ar sprijini ideea că, într-adevăr, conștiența este o parte vitală a liberului arbitru. Să ne gândim la o situație în care cineva îndeplinește o acțiune fără să-și dea seama ce face – acțiunea este realizată exclusiv prin activitatea nonconștientă a creierului persoanei respective. Am considera oare că aceasta e o manifestare a liberului arbitru? Majoritatea oamenilor ar răspunde *nu*. Dacă acțiunea ar fi nocivă, probabil că am considera chiar persoana respectivă răspunzătoare, dar am căuta acțiuni conștiente de dată recentă care ar fi putut-o determina să întreprindă acțiuni fără a fi conștientă – ca de pildă să bea prea mult sau pur și simplu să nu se pregătească adecvat pentru a-și cântări deciziile înainte de a le pune în aplicare.

Potrivit unor comentatori, experimentele lui Libet sunt împotriva liberului arbitru, scoțând în evidență măsura în care procesul de luare a deciziilor nu e conștient. De vreme ce în rândul filosofilor există un consens rezonabil în privința faptului că liberul arbitru presupune și luarea conștientă a deciziilor, aceasta pare să fie o cerință esențială a liberului arbitru. Pe de altă parte, totuși, pentru mulți observatori, coonștiența este o condiție necesară, dar nu suficientă. Dacă deciziile noastre – conștiente sau nu – sunt predeterminate înainte să le luăm, cum putem spune că deciziile noastre sunt libere? Această poziție, care susține că liberul arbitru și determinismul nu sunt compatibile, este cunoscută sub denumirea de incompatibilism. De exemplu, filosoful american Carl Ginet (născut în anul 1932) argumentează că, dacă evenimentele din trecut, prezent și viitor sunt determinate, atunci se poate considera că nu avem niciun fel de control asupra lor sau asupra

consecinţelor lor. Deciziile şi acţiunile noastre aparente sunt pur şi simplu parte a acestei succesiuni predeterminate. Pentru Ginet, asta elimină liberul arbitru.

Însă nu toată lumea consideră că determinismul este compatibil cu conceptul de liber arbitru. Compatibiliştii susţin, în esenţă, că eşti liber să decizi ce vrei, chiar dacă ceea ce decizi este sau poate fi determinat. Daniel Dennett, de exemplu, susţine cu argumente că, în vreme ce viitorul poate fi determinat din starea prezentului, lumea este atât de încâlcit complexă, încât de fapt nu avem cum şti ce ne va aduce viitorul. Putem identifica trăsăturile pe care Dennett le numeşte „aşteptări" şi suntem pe bună dreptate liberi să ducem la îndeplinire acţiuni care nu sunt în acord cu aceste aşteptări. Ar trebui să luăm în considerare felul cum arată deciziile şi acţiunile noastre comparativ cu aceste aşteptări şi nu cu viitorul nostru determinat teoretic, pe care, de fapt, nu-l putem cunoaşte. Acest lucru, susţine Dennett, este suficient pentru a confirma liberul arbitru.

Gazzaniga formulează şi o poziţie compatibilistă: „Suntem personal agenţi responsabili şi trebuie să fim consideraţi răspunzători pentru acţiunile noastre chiar dacă trăim într-o lume determinată".[17] Cinicul ar putea să interpreteze această perspectivă la modul: Nu ai control asupra acţiunilor tale, dar oricum o să dăm vina pe tine.

Unii gânditori resping ideea liberului arbitru. Filosoful scoţian David Hume (1711-1776) a descris-o pur şi simplu ca fiind o chestiune „verbală" caracterizată de o „falsă senzaţie sau de o pretinsă experienţă".[18] Filosoful german Arthur Schopenhauer (1788-1860) a scris că „oricine se consideră *a priori* perfect liber, chiar şi în acţiunile sale individuale, şi crede că în orice moment poate să înceapă un alt gen de viaţă... Dar *a posteriori*, prin experienţă, constată spre uimirea sa că nu este liber, ci e supus necesităţii, că, în ciuda tuturor hotărârilor sale, nu-şi schimbă purtarea şi că, de la începutul vieţii sale şi până la sfârşitul ei, trebuie să poarte chiar caracterul pe care el însuşi îl condamnă".[19]

Eu aş adăuga mai multe aspecte aici. Conceptul de liber arbitru – şi responsabilitate, care este o idee strâns legată – e util şi pe bună dreptate vital pentru menţinerea ordinii sociale, indiferent dacă liberul arbitru există sau nu. Aşa cum conştienţa există clar ca memă, tot aşa există şi liberul arbitru. Tentativele de dovedire a existenţei sale sau măcar de definire a ei pot deveni disperant de circulare, dar aproape toată lumea crede în această idee. Părţi substanţiale ale neocortexului nostru de nivel superior

sunt dedicate convingerii că facem alegeri libere şi că suntem răspunzători de acțiunile noastre. Într-un sens strict filosofic, acest lucru este adevărat sau măcar posibil, iar dacă că nu am avea asemenea credințe, societatea ar fi cu mult mai rea.

Mai mult de-atât, lumea nu este în mod necesar determinată. Am discutat mai înainte două perspective asupra mecanicii cuantice care se deosebesc din punctul de vedere al relației câmpurilor cuantice cu un observator. O interpretare populară a perspectivei observatorului conferă un rol conştienței: particulele nu-şi lămuresc ambiguitatea cuantică până în momentul în care sunt observate de un observator conştient. Mai există o scindare în filosofia evenimentelor cuantice care ne influențează discuția despre liberul arbitru, scindare care evoluează în jurul întrebării: Evenimentele cuantice sunt determinate sau aleatorii?

Cea mai obişnuită interpretare a unui eveniment cuantic spune că, atunci când funcția de undă care constituie o particulă „colapsează", localizarea particulei devine sigură. Pentru foarte multe asemenea evenimente va exista o distribuție predictibilă (motiv pentru care funcția de undă este considerată a fi o distribuție probabilistică), dar hotărârea privitoare la fiecare particulă care să sufere colapsul funcției de undă este aleatorie. Interpretarea în sens opus este deterministă: mai exact, există o variabilă ascunsă pe care nu suntem capabili s-o detectăm separat, dar a cărei valoare determină poziția particulei. Valoarea sau faza variabilei ascunse în momentul în care funcția de undă colapsează determină poziția particulei. Cei mai mulți fizicieni specialişti în mecanica cuantică par să agreeze ideea unei hotărâri aleatorii potrivit câmpului de probabilitate, dar ecuațiile mecanicii cuantice permit existența unei asemenea variabile ascunse.

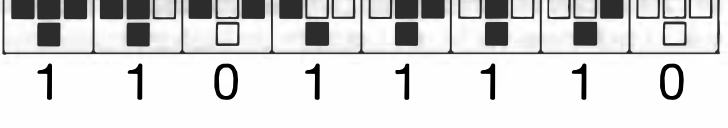
Prin urmare, s-ar putea ca lumea să nu fie deloc determinată. Potrivit interpretării mecanicii cuantice făcute prin prisma undei de probabilitate, la cel mai elementar nivel al realității există o sursă permanentă de incertitudine. Pe de altă parte însă, această observație nu rezolvă neapărat motivele de îngrijorare ale incompatibiliştilor. Este adevărat că, în această interpretare a mecanicii cuantice, lumea nu e determinată, dar conceptul nostru de liber arbitru ajunge până dincolo de deciziile şi acțiunile pur aleatorii. Cei mai mulți dintre incompatibilişti ar considera conceptul de liber arbitru incompatibil şi cu faptul că deciziile noastre sunt în esență accidentale. Liberul arbitru pare să presupună o luare intenționată a deciziilor.

Doctorul Wolfram propune un mod de rezolvare a acestei dileme. Cartea lui, *A New Kind of Science – Un nou gen de ştiinţă*, apărută în anul 2002, prezintă o perspectivă cuprinzătoare asupra ideii de automate celulare şi asupra rolului lor în toate aspectele vieţii noastre. Automatul celular este un mecanism în care valoarea celulelor de informaţie este recomputată continuu în funcţie de celulele din jurul ei. John von Neumann a creat o maşină teoretică autoreplicativă denumită constructor universal, care a fost, probabil, primul automat celular.

Doctorul Wolfram îşi ilustrează teza cu cel mai simplu automat celular cu putinţă, un grup de celule situate pe o linie unidimensională. În fiecare moment din timp, fiecare celulă poate să aibă două valori: neagră sau albă. Valoarea fiecărei celule este recomputată pentru fiecare ciclu. Valoarea unei celule pentru următorul ciclu este o funcţie care depinde şi de valoarea ei curentă, şi de valoarea celor doi vecini alăturaţi. Fiecare automat celular este caracterizat de o regulă care determină felul cum computăm dacă o celulă e neagră sau albă în următorul ciclu.

Să ne gândim, de exemplu, la ceea ce doctorul Wolfram numeşte regula 222.

Regula 222

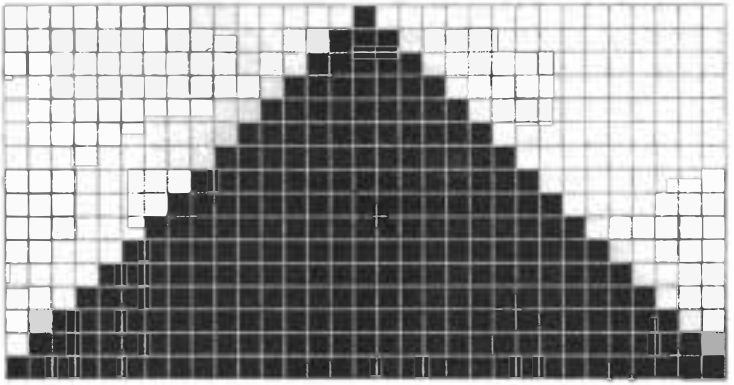


Cele opt posibile combinaţii ale valorilor celulei care este recomputată şi ale vecinilor ei din stânga şi din dreapta sunt prezentate pe rândul de sus. Noua valoare este reprezentată pe rândul de jos. Astfel, de exemplu, dacă celula este neagră şi cei doi vecini ai săi sunt de asemenea negri, celula rămâne neagră şi în următoarea generaţie (vezi subregula din extrema stângă a regulii 222). Dacă celula este albă, vecina ei din stânga e albă şi vecina din dreapta e neagră, atunci va deveni neagră în următoarea generaţie (vezi subregula regulii 222, a doua din dreapta).

Universul acestui automat celular simplu constă dintr-un singur rând de celule. Dacă începem cu o singură celulă neagră din mijloc şi arătăm evoluţia celulelor pe mai multe generaţii (în care fiecare rând, în sens descendent,

reprezintă o nouă generație de valori), rezultatul regulii 222 se prezintă după cum urmează:

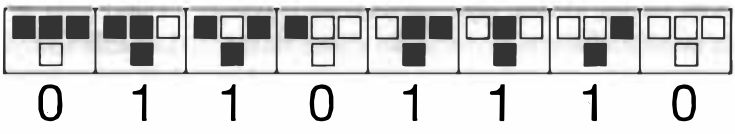
Regula 222



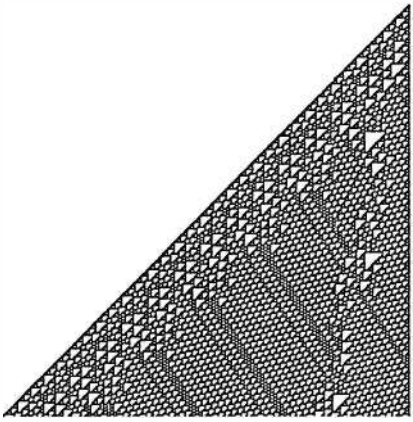
Automatul se bazează pe o regulă şi regula stabileşte dacă celula va fi neagră sau albă în funcție de cele opt modele posibile care există în generația curentă. Astfel, există $2^8 = 256$ de reguli posibile. Doctorul Wolfram a enumerat toate cele 256 astfel de automate posibile şi i-a atribuit fiecăruia un cod Wolfram de la 0 la 255. Interesant este că aceste 256 de maşini teoretice au proprietăți foarte diferite. Automatul din clasa I, cum îl numeşte Wolfram, de pildă regula 222, creează modele foarte predictibile. Dacă ar fi să te întreb ce valoare va avea celula din mijloc după o mie de miliarde de mii de miliarde de repetări ale regulii 222, ai putea să răspunzi cu uşurință: neagră.

Şi mai interesante, pe de altă parte, sunt automatele de clasa a IV-a, ilustrate de regula 110.

Regula 110



Generațiile multiple ale acestui automat arată astfel:



Aspectul interesant al automatului regulii 110 şi al automatelor din clasa a IV-a în general rezidă în faptul că rezultatele sunt complet impredictibile. Aceste rezultate trec şi cele mai stricte teste matematice ale caracterului aleatoriu, dar nu generează zgomot: sunt modele care se repetă, dar se repetă în moduri ciudate şi imprevizibile. Dacă te-aş întreba care este valoarea unei anumite celule după o mie de miliarde de mii de miliarde de repetări, nu ai putea să-mi răspunzi fără să treci această maşină prin toate generațiile respective. Soluția este clar determinată, deoarece avem de-a face cu o maşină deterministă foarte simplă, dar e complet impredictibilă în absența rulării propriu-zise a maşinii.

Teza primară a doctorului Wolfram este că lumea e un mare automat celular din clasa a IV-a. Motivul pentru care cartea lui se intitulează *A New Kind* of *Science – Un nou fel de ştiință* este acela că teza vine în opoziție cu majoritatea celorlalte legi ştiințifice. Dacă am avea un satelit care ar orbita în jurul Pământului, am putea anticipa locul unde se va afla peste cinci ani de-acum înainte fără să trebuiască să trecem prin fiecare moment al unui proces simulat, folosind legile relevante ale gravitației şi aflând unde se va afla satelitul în diferite momente din viitorul îndepărtat. Dar starea viitoare a automatului celular din clasa a IV-a nu poate fi anticipată fără simularea fiecărui pas. Dacă Universul este un gigantic automat celular, cum postulează doctorul Wolfram, nu ar putea să existe un computer suficient de

mare – pentru că fiecare computer ar fi un subset al Universului – pentru a putea să efectueze o asemenea simulare. Prin urmare, starea viitoare a Universului este complet imposibil de cunoscut, cu toate că e deterministă.

Astfel, cu toate că sunt determinate (deoarece corpul şi creierul nostru sunt parte a unui univers determinist), deciziile noastre sunt, totuşi, în mod inerent imprevizibile, deoarece trăim într-un automat de clasa a IV-a (şi suntem parte a lui). Nu putem anticipa viitorul unui automat de clasa a IV-a decât în cazul în care lăsăm acest viitor să se desfăşoare. Pentru doctorul Wolfram, este un fapt suficient pentru acceptarea liberului arbitru.

Nu trebuie să ne uităm la Univers pentru a vedea evenimente viitoare care sunt determinate dar imprevizibile. Niciunul dintre oamenii de ştiinţă care au lucrat la Watson nu poate să anticipeze ce va face acesta, deoarece programul este pur şi simplu prea variat şi prea complex, iar performanţa lui se bazează pe cunoştinţe mult prea extinse pentru a putea fi gestionate de om. Dacă credem că omul manifestă liber arbitru, va trebui să acceptăm că viitoarele versiuni ale lui Watson sau maşinile asemănătoare acestuia îl pot manifesta şi ele.

Propriul meu salt de credinţă este acela că omul are liber arbitru şi, în timp ce acţionez ca şi cum aşa este, sunt obligat să găsesc exemple din propriile mele decizii care să ilustreze asta. Să ne gândim la decizia de a scrie această carte – nu eu am luat-o. Dimpotrivă, ideea cărţii a luat decizia în locul meu. În general, mă văd captivul ideilor care par să se implanteze în neocortexul meu şi să preia comanda. Cum rămâne cu decizia de a mă căsători, pe care am luat-o (în colaborare cu cealaltă persoană) în urmă cu treizeci şi şase de ani? La momentul respectiv, am urmat programul obişnuit de a fi atras – şi de a urmări – o fată frumoasă. Pe urmă m-am îndrăgostit. Unde e liberul arbitru aici?

Dar cum rămâne cu micile decizii pe care le iau zi de zi – de exemplu, cuvintele pe care aleg să le scriu în cartea mea? Pornesc de la o foaie albă de hârtie. Nimeni nu-mi spune ce să fac. Nu există niciun redactor care să mă urmărească peste umăr. Alegerile mele depind *exclusiv* de mine. Sunt liber – *total liber* – să scriu *orice*...

Ah, *grok*...

Grok? În regulă, am comis-o – mi-am folosit în sfârşit liberul arbitru. Voiam să scriu cuvântul „vreau", dar am luat decizia de a scrie ceva total neaşteptat. Probabil că este prima oară când am reuşit să-mi exercit liberul arbitru pur.

Sau nu.

Ar trebui să fie evident că aceasta nu a fost o manifestare a liberului arbitru ci, dimpotrivă, o încercare de a ilustra o idee (şi, probabil, un simţ al umorului slăbuţ).

Deşi împărtăşesc convingerea lui Descartes că sunt conştient, nu sunt atât de convins în privinţa liberului arbitru. Este dificil să eviţi concluzia lui Schopenhauer că „poţi să faci ce vrei, dar, în orice moment al vieţii, poţi să-ţi doreşti un singur lucru şi numai unul şi nimic altceva în afara acelui lucru".[20] În orice caz, voi continua să mă port ca şi cum aş avea liber arbitru şi să cred în el, atâta vreme cât nu trebuie să explic de ce.

Identitatea

Un filosof a avut cândva următorul vis. Mai întâi i s-a arătat Aristotel şi filosoful i-a spus: „Poţi să-mi faci în cincisprezece minute o schiţă a întregii tale filosofii?" Spre surpriza filosofului, Aristotel i-a făcut o expunere excelentă, în care a comprimat un volum impresionant de material în numai cincisprezece minute. Dar, pe urmă, filosoful a avut o anumită obiecţie la care Aristotel nu a putut să răspundă. Pus în încurcătură, Aristotel a dispărut.

Pe urmă, i-a apărut Platon. S-a întâmplat iar acelaşi lucru, iar obiecţia filosofului la Platon a fost aceeaşi ca la Aristotel. Nici Platon nu a putut răspunde şi a dispărut.

Pe urmă i s-au arătat toţi filosofii celebri din istorie, unul câte unul, şi filosoful nostru i-a alungat pe toţi, pe rând, cu aceeaşi obiecţie. După ce a dispărut şi ultimul filosof, omul nostru şi-a spus: „Ştiu că dorm şi că visez toate aceste lucruri. Dar am descoperit respingerea universală pentru toate sistemele filosofice! Mâine dimineaţă, după ce mă voi trezi, probabil că o s-o uit, iar lumea o să piardă într-adevăr ceva!" Cu o voinţă de fier, filosoful s-a forţat să se trezească, s-a repezit la masă şi şi-a aşternut pe hârtie respingerea universală. Pe urmă a sărit înapoi în pat şi a răsuflat uşurat.

A doua zi dimineaţă, când s-a trezit, s-a dus la masă să vadă ce scrisese. Cuvintele erau: „Asta e părerea ta".

— Raymond Smullyan, citat de David Chalmers[21]

Ce mă întreb mai des decât dacă sunt conștient sau nu și dacă îmi exercit sau nu liberul arbitru este de ce sunt conștient de experiențele și deciziile acestei persoane anume care scrie cărți, este pasionată de drumeții și plimbări cu bicicleta, ia suplimente nutritive și așa mai departe. Un răspuns evident ar fi „Pentru că ăsta eşti".

Probabil că acest schimb nu este mai tautologic decât răspunsurile mele de mai sus la întrebările legate de conștiință și de liberul arbitru. Dar, de fapt, am un răspuns mai bun la întrebarea de ce conștiența mea este asociată cu această persoană anume: Pentru că acesta m-am creat să fiu.

Un aforism cunoscut spune: „Eşti ceea ce mănânci". Și mai adevărat este să spui: „Eşti ceea ce gândeşti". Așa cum am discutat, toate structurile ierarhice din neocortexul meu care îmi definesc personalitatea, abilitățile și cunoștințele sunt rezultatul propriilor mele gânduri și experiențe.

Oamenii cu care aleg să interacționez, ideile și proiectele în care aleg să mă angajez sunt, în totalitatea lor, determinanții primari ai persoanei care am devenit. Din acest punct de vedere, ceea ce mănânc reflectă de asemenea deciziile luate de neocortexul meu. Acceptând pentru moment latura pozitivă a dualității liberului arbitru, propriile mele decizii s-au concretizat în cel care sunt.

Indiferent de felul cum am ajuns să fim cei care suntem, fiecare dintre noi are dorința ca identitatea lui să se mențină. Dacă nu aveai dorința de a supraviețui, nu te-ai afla unde eşti și nu ai citi această carte. Orice ființă are acest scop – este principalul determinant al evoluției. Subiectul identității e, probabil, și mai greu de definit decât acela al conștienței sau al liberului arbitru, dar este – demonstrabil – cel mai important. La urma urmelor, trebuie să știm ce suntem dacă vrem să încercăm să ne conservăm existența.

Gândeşte-te la acest experiment mintal: te afli cândva în viitor și ai la dispoziție tehnologii mai avansate decât cele de azi. Când dormi, un anumit grup îți scanează creierul și copiază toate detaliile ieşite în evidență. Echipa poate face acest lucru cu mașini de scanat de mărimea elementelor figurate ale sângelui, mașini care circulă prin capilarele din creierul tău, sau prin intermediul unei alte tehnologii neinvazive, însă are toată informația din creierul tău la un moment dat. De asemenea, citeşte și înregistrează orice detaliu fizic al corpului tău care ar putea să se reflecte în starea ta de spirit, ca de pildă sistemul endocrin. Imortalizează acest „fișier al minții" într-un corp nonbiologic care seamănă, se mișcă la fel ca tine și are subtilitatea și suplețea necesare pentru a trece drept tine. Dimineața, eşti informat asupra

acestui transfer şi priveşti (poate fără să fii observat) clona minţii tale, pe care o vom denumit Tu 2. Acest Tu 2 vorbeşte despre viaţa lui ca şi cum ar fi tu şi dezvăluie cum a descoperit chiar în acea dimineaţă că a primit o nouă versiune mai durabilă a corpului, 2.0. „Hei, îmi şi place acest nou corp!", afirmă Tu 2.

Prima întrebare la care trebuie să te gândeşti este: Tu 2 e conştient? Păi, în mod evident, pare să fie. Tu 2 trece testul pe care l-am conceput anterior, în sensul că prezintă indiciile subtile că ar fi o persoană înzestrată cu sensibilitate şi conştienţă. Dacă tu eşti conştient, înseamnă că şi Tu 2 este conştient.

Prin urmare, dacă ai dispărea, să spunem, nimeni n-ar băga de seamă. Tu 2 ar putea să se plimbe de colo-colo pretinzând că este tu. Toţi prietenii tăi şi persoanele iubite de tine s-ar mulţumi cu această situaţie şi probabil că ar fi încântate că de-acum ai un corp şi un substrat mintal mai durabile decât înainte. Poate că prietenii tăi mai înclinaţi către filosofie şi-ar exprima îngrijorarea, dar, în cea mai mare parte, toată lumea ar fi fericită, inclusiv tu – sau cel puţin persoana care afirmă în mod convingător că ar fi tu.

Deci nu mai avem nevoie nici de vechiul corp şi nici de vechiul creier, nu? E-n regulă dacă ne descotorosim de ele?

Probabil că nu vei dori să accepţi aşa ceva. Am spus că scanarea a fost neinvazivă, aşa încât tu eşti în continuare prezent şi conştient. Mai mult de-atât, simţul identităţii continuă să te însoţească, nu s-a dus cu Tu 2, deşi Tu 2 crede că e o continuare a ta. Este posibil chiar ca Tu 2 să nici nu fie conştient că există sau că ai existat vreodată. De fapt, nu ai fi fost nici tu conştient de existenţa lui Tu 2 dacă nu ţi-am fi spus noi.

Concluzia noastră? Tu 2 este conştient, dar este o altă persoană, distinctă de tine – Tu 2 are o identitate diferită. Îţi este extrem de similar, cu mult mai mult decât o simplă clonă genetică, pentru că îţi împărtăşeşte inclusiv toate modelele şi conexiunile neocorticale. Sau poate că ar trebui să spun că a împărtăşit aceste modele în momentul în care a fost creat. Din acel moment, voi doi aţi pornit fiecare pe calea lui din punct de vedere neocortical. Tu eşti în continuare prezent. Nu ai aceleaşi experienţe ca Tu 2. Concluzia: Tu 2 nu este tu.

În regulă, toate bune şi frumoase până aici. Acum, să ne gândim la un alt experiment mintal – unul care este, cred eu, mai realist în termeni de surprize ale viitorului. Eşti supus unei proceduri de înlocuire a unei foarte mici porţiuni din creierul tău cu o unitate nonbiologică. Eşti convins că nu e

ceva nociv şi că au fost comunicate numeroase beneficii legate de această operație.

Nu e ceva extrem de deplasat, deoarece această procedură le este aplicată în mod curent persoanelor care suferă de dizabilități neurologice şi senzoriale, cum ar fi implanturile neuronale pentru boala Parkinson şi implanturile cohleare pentru surzi. În aceste cazuri, dispozitivul computerizat este plasat în interiorul corpului tău, însă în afara creierului, dar e conectat la creier (sau, în cazul implanturilor cohleare, la nervul auditiv). Din punctul meu de vedere, faptul că acest computer este plasat, practic, în afara creierului propriu-zis nu reprezintă ceva semnificativ sub aspect filosofic: efectiv amplificăm capacitățile creierului şi înlocuim cu un dispozitiv computerizat acele funcții ale sale care nu mai sunt adecvate. În anii 2030, când dispozitivele computerizate inteligente vor avea mărimea elementelor figurate ale sângelui (şi reține că globulele albe sunt suficient de inteligente ca să recunoască şi să combată agenții patogeni), ele vor fi introduse noninvaziv, fără să fie necesară o intervenție chirurgicală.

Revenind la acest scenariu din viitor, ți s-a aplicat procedura şi, cum ți s-a promis, ea dă rezultate bune – anumite capacități ale tale s-au îmbunătățit. (Ai o memorie mai bună, poate.) Deci mai eşti tot tu sau nu? Prietenii tăi cu siguranță cred că eşti. Şi tu crezi la fel. Nu există niciun argument solid că ai fi devenit brusc o altă persoană. În mod evident, ai fost supus procedurii de efectuare a unui anumit schimb, dar eşti în continuare tu, cel dintotdeauna. Identitatea ta nu s-a modificat. Nu se pune problema că o conştiență a altcuiva ar fi pus brusc stăpânire pe corpul tău.

În regulă. Aşadar, încurajat de aceste rezultate, decizi acum să mai urmezi o procedură, de această dată interesând o altă regiune a creierului tău. Rezultatul este acelaşi: resimți o anumită ameliorare sub aspectul capacității, dar eşti în continuare tot tu.

Ar trebui să fie evident unde bat cu asta. Continui să optezi pentru alte şi alte proceduri, iar încrederea ta în acest proces nu face decât să crească, până când, în cele din urmă, ți-ai schimbat până şi ultima parte din creier. De fiecare dată, procedura a fost realizată cu cea mai mare grijă, pentru păstrarea tuturor modelelor şi conexiunilor tale neocorticale, astfel încât nu ți-ai pierdut nimic din personalitate, abilități şi memorie. Nu a existat niciodată un Tu şi un Tu 2; ai fost doar Tu. Nimeni nu observă că încetezi să mai exişti – nici măcar tu însuți. Şi pe bună dreptate – uită-te la tine!

Concluzia noastră: Tu continui să exişti. Nu este loc pentru nicio dilemă aici. Totul este în regulă.

Cu următoarea excepție: după procesul treptat de înlocuire, eşti pe deplin echivalent cu Tu 2 din anteriorul experiment mintal (pe care îl voi numi *scenariu de scanare şi imortalizare*). Tu, cel de după scenariul de înlocuire, ai toate modelele şi conexiunile neocorticale pe care le aveai inițial, doar că într-un substrat nonbiologic, ceea ce este adevărat şi pentru Tu 2 din scenariul scanării şi imortalizării. Tu, cel de după aplicarea scenariului înlocuirii treptate, ai unele capacități suplimentare şi o mai mare durabilitate decât aveai înaintea procesului, însă acest lucru e la fel de adevărat şi pentru Tu 2 din procesul de scanare şi imortalizare.

Dar am concluzionat că Tu 2 nu este una şi aceeaşi persoană cu tine. Iar dacă tu, după procesul de înlocuire treptată, eşti absolut echivalent cu Tu 2 de după procesul de scanare şi imortalizare, înseamnă că tu cel de după procesul de înlocuire treptată trebuie să nu mai fii tot tu.

Situația aceasta ne contrazice însă concluzia anterioară. Procesul de înlocuire treptată constă în mai mulți paşi. Fiecare din aceşti paşi a părut să conserve identitatea, la fel cum am concluziona astăzi că o persoană care suferă de boala Parkinson are aceeaşi identitate şi după instalarea unui implant neuronal.[22]

Exact acest gen de dilemă filosofică îi determină pe unii să concluzioneze că scenariile de înlocuire nu vor avea loc niciodată (cu toate că ele se petrec deja). Dar să ne gândim la următorul lucru: în mod natural, suntem supuşi unui proces de înlocuire treptată pe parcursul întregii noastre vieți. Cele mai multe celule din corpul nostru sunt înlocuite continuu. (Tocmai ți s-au înlocuit 100 de milioane de celule cât timp ai citit ultima propoziție.) Celulele mucoasei interne a intestinului subțire se schimbă în aproximativ o săptămână, la fel ca mucoasa protectoare a stomacului. Durata de viață a globulelor albe ale sângelui ajunge de la câteva zile la câteva luni, în funcție de tip. Plachetele sanguine trăiesc în jur de nouă zile.

Neuronii persistă, însă organitele şi moleculele care le compun se schimbă în decurs de o lună.[23] Timpul de înjumătățire a microtubulilor din neuron este de circa zece minute; filamentele de actină din dendrite trăiesc în jur de patruzeci de secunde; proteinele care asigură energia sinapselor sunt înlocuite la fiecare oră; receptorii NMDA din sinapse au o viață relativ lungă, de circa cinci zile.

Prin urmare, eşti înlocuit aproape complet în decurs de câteva luni, fapt comparabil cu scenariul înlocuirii treptate pe care l-am descris anterior. Eşti aceeaşi persoană care erai în urmă cu câteva luni? Sigur că există unele

diferențe. Poate că ai învățat câteva lucruri noi. Dar presupui că identitatea ta se păstrează, că nu eşti distrus şi recreat continuu.

Gândeşte-te la un râu cum este cel care trece pe lângă biroul meu. Acest râu, căruia oamenii îi spun Charles, este acelaşi cu râul pe care l-am văzut ieri? Mai întâi, hai să reflectăm la ceea ce este un râu. Dicționarul îl definește ca „șuvoi natural, mare, de apă curgătoare". Potrivit acestei definiții, râul pe care îl privesc este complet diferit de cel care era ieri. Fiecare dintre moleculele sale de apă s-a schimbat, proces care se petrece foarte repede. Filosoful grec Diogene Laertius a scris în secolul trei al erei noastre că „nu poți să calci de două ori în apa aceluiaşi râu".

Însă nu aşa privim în general râurile. Oamenilor le place să le privească deoarece ele sunt simboluri ale continuității şi stabilității. Potrivit concepției generale, râul Charles pe care l-am privit ieri este acelaşi râu pe care îl privesc şi azi. Viața noastră este aproape aceeaşi. În esență, nu suntem materia din care este alcătuit corpul şi creierul nostru. În esență, aceste particule curg prin noi la fel cum curg particulele de apă prin râu. Suntem un model care se schimbă încetul cu încetul, dar are stabilitate şi continuitate, chiar dacă materia care alcătuieşte modelul se schimbă rapid.

Introducerea treptată a sistemelor nonbiologice în corpul şi creierul nostru nu va fi altceva decât un nou exemplu de înnoire continuă a părților care ne alcătuiesc. Asta nu ne va altera continuitatea identității mai mult decât înlocuirea naturală a celulelor noastre biologice. În cea mai mare parte, ne-am externalizat memoria istorică, intelectuală, socială şi personală în dispozitive electronice şi în nor. Dispozitivele cu care interacționăm pentru accesarea acestei memorii poate că nu se află încă în interiorul corpului şi creierului nostru, dar, pe măsură ce devin din ce în ce mai mici (şi ne micşorăm tehnologia cu o rată de circa 100 în volum pe deceniu), îşi vor croi drumul către interiorul nostru. În orice caz, le vom pune într-un loc util – şi nici nu le vom pierde în acest fel. Dacă oamenii vor opta într-adevăr pentru plasarea unor dispozitive microscopice în interiorul corpului lor, acest lucru va fi în regulă, deoarece vor exista şi alte moduri de accesare a inteligenței omniprezente a norului.

Dar să revenim la dilema pe care am propus-o mai înainte. Tu, după o perioadă de înlocuire treptată, eşti echivalent cu Tu 2 din scenariul scanării şi imortalizării, dar am decis că Tu 2 din acel scenariu nu are aceeaşi identitate ca tine. Prin urmare, cu ce ne-am ales?

Ne-am ales cu înțelegerea unei capacități pe care sistemele nonbiologice o au, iar sistemele biologice n-o au: capacitatea de a fi copiate pentru

siguranță, salvate şi recreate. Facem în mod curent acest lucru cu dispozitivele noastre. Când ne cumpărăm un *smartphone* nou, copiem în el toate fişierele, astfel că el are, în mare, cam aceeaşi personalitate, aceleaşi abilități şi amintiri ca vechiul *smartphone*. Poate că are şi o serie de capacități noi, dar conținutul vechiului telefon ne însoțeşte mai departe. În mod similar, un program cum este Watson are în mod cert o copie de siguranță. Dacă hardware-ul lui Watson ar fi distrus mâine, Watson ar putea fi recreat cu uşurință din fişierele de siguranță stocate în nor.

Aceasta reprezintă o capacitate a lumii nonbiologice care nu există în lumea biologică. Este un avantaj, nu o limitare, motiv pentru care astăzi suntem foarte dornici să continuăm să ne încărcăm memoriile în nor. În mod cert vom continua în această direcție, pe măsură ce sistemele nonbiologice vor căpăta tot mai multe capacități ale creierului nostru biologic.

Rezolvarea pe care o dau eu acestei dileme este următoarea: Nu e adevărat că Tu 2 nu eşti tot Tu – eşti Tu. Doar că acum sunt doi Tu. Nu e chiar atât de rău – dacă crezi că eşti bun, atunci doi de Tu înseamnă şi mai bine.

Ce cred că se va întâmpla de fapt este că vom continua pe calea scenariului înlocuirii treptate şi al augmentării până când, în cele din urmă, cea mai mare parte a gândirii noastre se va realiza în nor. Saltul meu de credință în privința identității este că aceasta se conservă prin continuitatea modelului de informații ce ne face oamenii care suntem. Continuitatea permite schimbarea permanentă, astfel încât, cu toate că sunt întrucâtva diferit de cel care eram ieri, am, totuşi, aceeaşi identitate. Pe de altă parte, continuitatea modelului care constituie identitatea mea nu e dependentă de substrat. Substraturile biologice sunt minunate – ne-au dus foarte departe –, însă noi creăm acum un substrat mai capabil şi mai durabil din motive cât se poate de pozitive.

Capitolul 10

LEGEA ACCELERĂRII VENITURILOR APLICATĂ CREIERULUI

Cu toate că omul ar trebui să rămână, din anumite puncte de vedere, ființa superioară, oare acest lucru nu este în acord cu practica naturii, care va accepta superioritatea, sub anumite aspecte, a unor animale care au fost de mult depășite? Oare nu le-a permis furnicii și albinei să-și păstreze superioritatea față de om prin organizarea comunităților lor și prin aranjamentele sociale, păsării prin străbaterea văzduhului, peștelui prin înot, calului prin forță și iuțeală și câinelui prin simțul sacrificiului de sine?

— Samuel Butler, 1871

A fost o vreme când Pământul era, după toate aparențele, lipsit de orice formă de viață animală și vegetală și când, potrivit opiniei celor mai buni filosofi ai noștri, era doar o sferă încinsă, cu o crustă care se răcea treptat. Dacă ar fi existat vreun om pe vremea când Pământul era în acea stare și ar fi putut să-l vadă ca pe un fel de altă lume cu care nu avea nimic în comun, dacă în același timp ar fi fost neștiutor în știința fizicii, n-ar fi spus el, oare, că este imposibil ca ființe capabile să posede ceva asemănător conștienței să evolueze din cenușa pe care o vedea? N-ar fi negat că aceasta conținea vreun potențial de conștiență? Și totuși, în decursul timpului, a apărut și conștiența. Nu este posibil, atunci, să existe chiar și acum noi canale pentru conștiență, deși nu putem să le discernem, în prezent, indiciile?

— Samuel Butler, 1871

Când reflectăm la multitudinea fazelor vieții și conștienței care au evoluat deja, ar însemna să ne pripim dacă am spune că nu se mai pot dezvolta și altele și că viața animală este sfârșitul tuturor lucrurilor. A fost o vreme în care focul era sfârșitul tuturor lucrurilor; o alta când piatra și apa erau sfârșitul.

— Samuel Butler, 1871

> Nu există nicio probă împotriva dezvoltării în ultimă instanță a conștienței în faptul că acum mașinile au prea puțină conștiență. O moluscă nu are prea multă conștiență. Reflectați la extraordinarul progres pe care l-au făcut mașinile în decursul ultimelor câteva sute de ani și observați cum progresează, încetul cu încetul, regnul animal și vegetal. Mașinile superior organizate nu sunt atât creaturi ale zilei de ieri, cât ale ultimelor cinci minute, ca să spun așa, raportat la trecut. Să presupunem de dragul argumentației că ființe conștiente există de aproximativ douăzeci de milioane de ani: vedeți ce progrese au făcut mașinile în ultima mie de ani! Nu poate lumea să mai dureze încă douăzeci de milioane de ani? Și atunci ce anume n-ar putea ele să devină în cele din urmă?
>
> **— Samuel Butler, 1871**

Teza mea fundamentală, pe care o numesc Legea accelerării veniturilor, este că mărimile fundamentale ale tehnologiei informației urmează traiectorii predictibile și exponențiale, dezmințind prejudecata că „nu poți prezice viitorul". Sunt încă multe lucruri – care proiect, companie sau standard tehnic vor prevala pe piață, când va veni pacea în Orientul Mijlociu – care rămân de neștiut, dar raportul preț/performanță și capacitatea informației s-au dovedit, totuși, remarcabil de predictibile. În mod surprinzător, aceste tendințe rămân neperturbate de condiții cum ar fi războiul sau pacea, prosperitatea sau recesiunea.

Un motiv principal pentru care evoluția a creat creiere l-a reprezentat prezicerea viitorului. Când unul dintre strămoșii noștri mergea prin savană, cu mii de ani în urmă, probabil că a observat că un animal înainta către drumul pe care îl urma și el. Putea să-și dea seama că, dacă rămâneau pe aceleași direcții, drumurile lor urmau să se intersecteze. Pe baza acestei constatări, el a decis să se îndrepte într-o altă direcție, iar anticiparea s-a dovedit prețioasă pentru supraviețuire.

Însă acești predictori incluși ai viitorului sunt liniari, nu exponențiali, calitate care își are originea în organizarea liniară a neocortexului. Să ne amintim că neocortexul face în permanență predicții – ce literă și ce cuvânt vom vedea imediat, pe cine ne așteptăm să vedem după primul colț de drum și așa mai departe. Neocortexul este organizat în succesiuni liniare de pași pentru fiecare model, ceea ce înseamnă că gândirea exponențială nu este ceva natural pentru noi. Și cerebelul folosește predicțiile liniare. Când ne ajută să prindem o minge din zbor, face o predicție liniară a locului în care

se va afla mingea în câmpul nostru vizual şi a celui în care ar trebui să fie mâna în mănuşă, în câmpul vizual, pentru a o prinde.

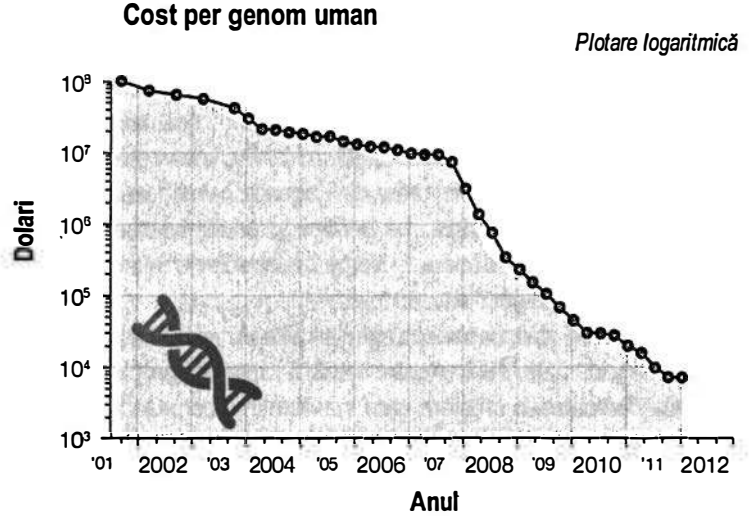
După cum am subliniat, există diferențe dramatice între progresia liniară şi cea exponențială (patruzeci de paşi liniari înseamnă tot patruzeci, pe când exponențiali înseamnă o mie de miliarde), ceea ce explică motivul pentru care predicțiile mele izvorâte din Legea accelerării veniturilor le par surprinzătoare, la prima vedere, multor observatori. Trebuie să ne antrenăm să gândim exponențial. Când este vorba de tehnologia informației, acesta este modul corect de gândire.

Exemplul cel mai pur pentru Legea accelerării veniturilor este creşterea perfect neaccidentată, dublu exponențială, a raportului preț/performanță în domeniul computației, creştere care s-a menținut constantă timp de 110 ani, deşi lumea a cunoscut două Războaie Mondiale, Marea Recesiune şi Războiul Rece, colapsul Uniunii Sovietice, reemergența Chinei, recenta criză financiară şi toate celelalte evenimente notabile de secol nouăsprezece, douăzeci şi început de secol douăzeci şi unu. Unii se referă la acest fenomen ca la „Legea lui Moore", însă ideea e greşită. Legea lui Moore – care afirmă că la fiecare doi ani se dublează numărul de componente de pe un circuit integrat care funcționează mai rapid pentru că şi componentele sunt mai mici – este doar una dintre multele paradigme. De fapt, creşterea exponențială a raportului preț/performanță în computație a fost produsă de cea de-a cincea paradigmă, nu de prima.

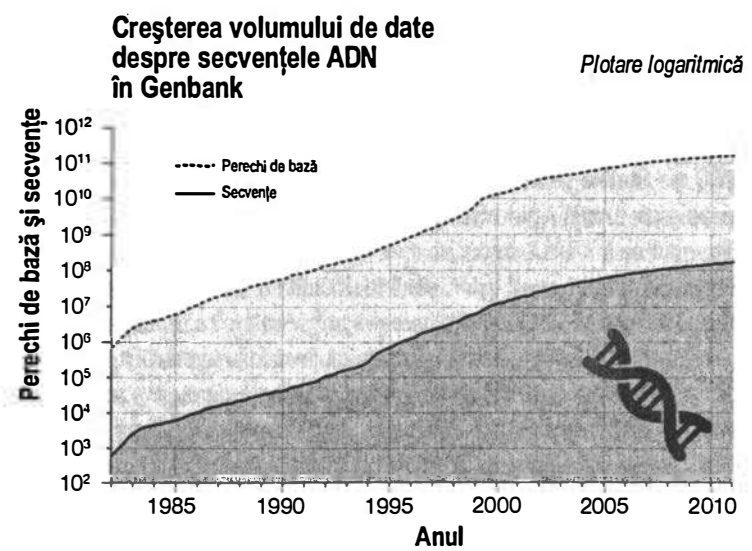
Creşterea exponențială a computației a debutat la recensământul din 1890 din Statele Unite (primul care s-a realizat automat), folosindu-se prima paradigmă a calculului electromecanic, cu zeci de ani înainte ca Gordon Moore să se fi născut. În cartea mea *The Singularity Is Near – Singularitatea e aproape*, prezint acest grafic până în anul 2002, iar aici îl actualizez până în 2009 (vezi graficul intitulat *Creşterea exponențială a computației timp de 110 ani*). Traiectoria uşor predictibilă a continuat, în ciuda recentului recul economic.

Computația este cel mai important exemplu de lege a accelerării veniturilor, grație volumului de date pe care îl avem în acest domeniu, al omniprezenței computației şi al rolului său cheie în revoluționarea, în ultimă instanță, a tot ceea ce ne este drag şi ne interesează. Dar nu este nici pe departe unicul exemplu. Din momentul în care tehnologia devine tehnologie a informației, ea începe să se supună Legii accelerării veniturilor.

Biomedicina devine cel mai important nou domeniu al tehnologiei şi industriei care e transformat în acest fel. Progresul în medicină s-a bazat,



Costul secvențializării unui genom de mărime umană[1]



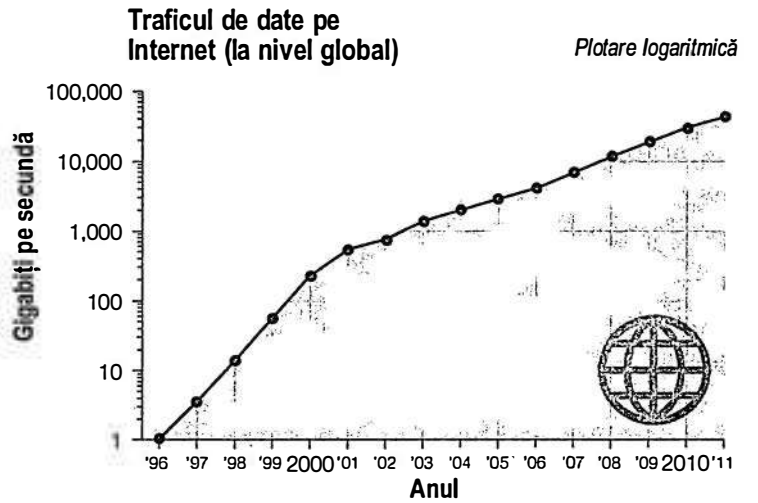
Volumul de date genetice secvențializate în lume în fiecare an[2]

din perspectivă istorică, pe descoperiri accidentale, astfel încât avansul în cursul epocii anterioare a fost liniar, nu exponențial. Cu toate acestea, a fost un fapt benefic: speranța de viață a crescut de la douăzeci și trei de ani în urmă cu un mileniu, la treizeci și șapte de ani în urmă cu două secole și la aproape optzeci de ani astăzi. Odată cu culegerea software-ului vieții – genomul –, medicina și biologia umană au devenit tehnologii ale informației. Însuși Proiectul genomului uman a fost perfect exponențial, volumul de date genetice dublându-se, iar costul la perechea de baze reducându-se cu jumătate în fiecare an de la înființarea proiectului, în 1990.[3] (Toate graficele din acest capitol au fost actualizate după publicarea cărții mele *The Singularity Is Near – Singularitatea este aproape*.)

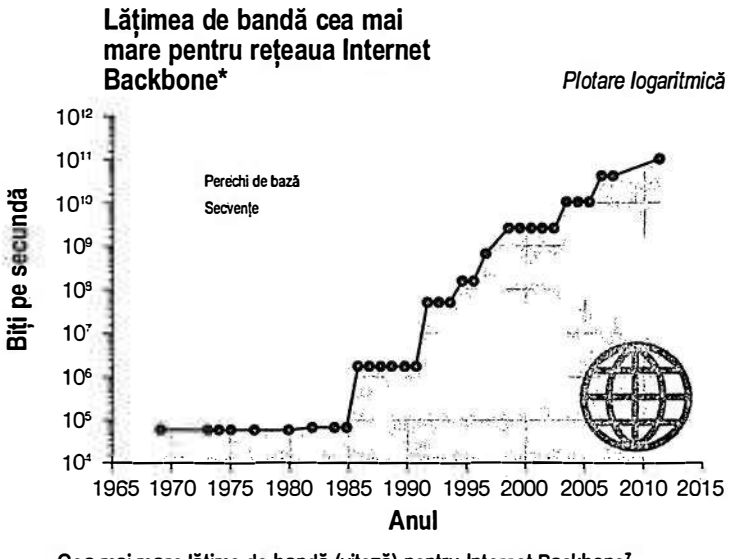
Actualmente, avem capacitatea de a concepe intervenții biomedicale pe computere și de a le testa pe simulatoare biologice, ale căror proporții și precizie se dublează și ele de la an la an. Putem de asemenea să ne actualizăm propriul software desuet: interferența ARN poate să deconecteze genele și noi forme de terapie genetică pot să adauge gene noi nu doar unui nounăscut, ci și unui individ matur. Progresul tehnologiilor genetice afectează și procesul de prelucrare a creierului prin metodele retroingineriei, prin faptul că un aspect important al său este înțelegerea modului în care genele controlează unele funcții ale creierului, cum ar fi crearea de noi conexiuni care să reflecte cunoștințele corticale adăugate recent. Există numeroase alte manifestări ale acestei întrepătrunderi a biologiei cu tehnologia informației, pe măsură ce depășim secvențierea genomului în direcția sintetizării acestuia.

O altă tehnologie a informației care a cunoscut o creștere exponențială neaccidentată o reprezintă capacitatea noastră de a comunica între noi și de a transmite vaste baze de cunoștințe umane. Există numeroase moduri de măsurare a acestui fenomen. Legea lui Cooper, care spune că totalul capacității comunicațiilor *wireless* pe un anumit interval al spectrului radio se dublează la fiecare trei luni, a rămas valabilă din vremea în care Guglielmo Marconi a folosit telegraful fără fir pentru transmisiuni în codul Morse în anul 1897 și până la tehnologiile comunicațiilor 4G din zilele noastre.[4] Potrivit Legii lui Cooper, volumul de informație care poate fi transmis pe un anumit interval al spectrului radio s-a dublat la fiecare doi ani și jumătate, timp de mai bine de un secol. Un alt exemplu îl constituie numărul de biți pe secundă transmiși pe Internet, care se dublează la fiecare 1,4 ani.[5]

Motivul pentru care am început să fiu interesat de anticiparea anumitor aspecte ale tehnologiei este că, în urmă cu aproximativ treizeci de ani,



Lățimea de bandă internațională (interțări) alocată Internetului pe plan mondial[6]



Cea mai mare lățime de bandă (viteză) pentru Internet Backbone[7]

*** Internet Backbone** este termenul prin care se face referire la traseele mari de date dintre rețelele mari, interconectate strategic, şi ruterele centrale de pe Internet. Aceste trasee de date sunt găzduite de centre de rețea comerciale, academice, guvernamentale, precum şi de alte centre de rețea de mare capacitate etc. (n.trad. – sursa: Wikipedia)

mi-am dat seama că soluția pentru a avea succes ca inventator (profesie pe care am îmbrățișat-o pe când aveam cinci ani) este sincronizarea. Cele mai multe invenții și cei mai mulți inventatori au eșuat nu pentru că născocirile respective nu ar fi funcționat, ci pentru că sincronizarea lor a fost greșită, ele apărând fie înainte de a exista toți factorii potențatori, fie prea târziu, ratând astfel fereastra de oportunitate.

Fiind inginer, în urmă cu aproximativ trei decenii am început să adun date privitoare la dimensiunile tehnologiilor din diverse domenii. Când am început această activitate, nu m-am așteptat să-mi dea o imagine clară, însă am sperat să îmi asigure o anumită orientare și să-mi permită să fac predicții în cunoștință de cauză. Scopul meu era – și este în continuare – să-mi sincronizez eforturile din domeniul tehnologic propriu, astfel încât acestea să fie adecvate pentru lumea existentă în momentul în care îmi închei un proiect – lume care, mi-am dat seama, este foarte diferită de cea existentă în momentul de debut al proiectului.

Să ne gândim acum cât de mult și cât de rapid s-a schimbat lumea în ultimul timp. În urmă cu numai câțiva ani, oamenii nu foloseau rețelele sociale (Facebook, de exemplu, a fost fondat în anul 2004 și avea 901 milioane de utilizatori activi la sfârșitul lunii martie 2012)[8], wiki, bloguri sau tweet. În anii 1990, cea mai mare parte din populația lumii nu folosea motoare de căutare sau telefoane celulare. Imaginează-ți lumea fără ele. Pare ceva de domeniul istoriei antice, dar se întâmpla nu cu mult timp în urmă. Lumea se va schimba și mai dramatic în viitorul apropiat.

În cursul investigațiilor mele, am făcut o descoperire uluitoare: dacă o tehnologie este una a informației, mărimile fundamentale ale raportului preț/performanță și capacitate (raportată la unitatea de timp, cost sau la altă resursă) urmează traiectorii precis exponențiale.

Aceste traiectorii depășesc paradigmele specifice pe care se bazează (cum ar fi Legea lui Moore). Însă atunci când o paradigmă rămâne fără elan (de exemplu, momentul când inginerii nu au mai putut să reducă mărimea sau costurile tuburilor cu vid în anii 1950), se generează presiune asupra cercetării pentru crearea următoarei paradigme și astfel debutează o nouă curbă S a progresului.

Partea exponențială a curbei S care urmează și vizează noua paradigmă continuă apoi curba permanent exponențială a măsurilor tehnologiei informației. Astfel, computarea bazată pe tuburile cu vid din anii 1950 a lăsat cale liberă pentru tranzistori în anii 1960 și pe urmă pentru circuitele integrate

şi Legea lui Moore la sfârşitul anilor 1960 şi după aceea. La rândul ei, Legea lui Moore va lăsa cale liberă computării tridimensionale, ale cărei exemple timpurii sunt deja implementate. Motivul pentru care tehnologiile informației sunt capabile să transceandă în mod sistematic limitele oricărei paradigme este acela că resursele necesare computației, recuperării sau transmiterii unui bit de informație sunt din ce în ce mai mici.

Ne-am putea întreba: Chiar există limite fundamentale pentru abilitatea noastră de computație şi de transmitere a informației indiferent de paradigmă? Răspunsul este da, pe baza felului în care înțelegem actualmente fizica computației. Pe de altă parte, aceste limite nu sunt prea limitative. În ultimă instanță, ne putem extinde inteligența de miliarde de ori pe baza computației moleculare. După calculele mele, vom atinge aceste limite la sfârşitul acestui secol.

Este important de subliniat faptul că nu orice fenomen exponențial reprezintă un exemplu pentru Legea accelerării veniturilor. Unii observatori interpretează greşit această lege, citând tendințe exponențiale care nu se bazează pe informație. De pildă, ei invocă faptul că aparatele de ras pentru bărbați au ajuns de la o lamă la două sau la patru şi apoi se întreabă unde sunt aparatele de ras cu opt lame. Numai că aparatele de ras nu înseamnă (deocamdată) tehnologie a informației.

În cartea mea *Singularity Is Near – Singularitatea este aproape*, am făcut o examinare teoretică, incluzând (în anexa cărții) o tratare matematică a motivului pentru care Legea accelerării veniturilor este remarcabil de predictibilă. În esență, folosim întotdeauna vârful ultimei tehnologii pentru a o crea pe următoarea. Tehnologiile se construiesc singure într-o manieră exponențială şi acest fenomen e uşor de măsurat dacă implică tehnologia informației. În anul 1990, foloseam computere şi alte instrumente ale erei respective pentru crearea computerelor anului 1991; în anul 2012, utilizăm actualele instrumente ale informației pentru a crea maşinile anilor 2013 şi 2014. Vorbind într-un sens mai larg, această accelerare şi dezvoltare exponențială se aplică oricărui proces în cadrul căruia evoluează modele de informație. Prin urmare, observăm accelerarea ritmului evoluției biologice, dar şi o accelerare similară (şi mult mai rapidă) a evoluției tehnologice, care este ea însăşi o ramificație a evoluției biologice.

Actualmente, mă aflu în posesia unui palmares public de peste un sfert de secol de predicții bazate pe Legea accelerării veniturilor, începând cu cele prezentate în cartea mea intitulată *The Age of Intelligent Machines – Epoca*

mașinilor inteligente, pe care am scris-o la mijlocul anilor 1980. Printre exemplele de predicții precise din această carte se numără: apariția spre sfârșitul anilor 1990 a unei vaste rețele mondiale de comunicații, care va conecta oamenii din toate colțurile lumii unii cu alții și cu toate cunoștințele umane; apariția unui puternic val de democratizare generat de aceste rețele descentralizate de comunicații care au măturat Uniunea Sovietică; înfrângerea campionului mondial la șah înainte de 1998 și multe altele.

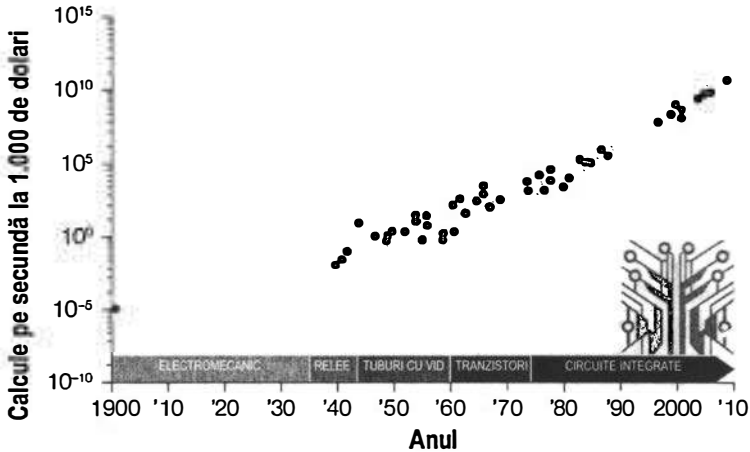
Am descris pe larg Legea accelerării veniturilor, așa cum este ea aplicată computației, în *The Age of Intelligent Machines – Epoca mașinilor inteligente*, carte în care am oferit, pentru un secol întreg, și datele ce indică progresul dublu exponențial al raportului preț/performanță al computației până în anul 1998. În continuare, le completez până în anul 2009.

Recent, am realizat o trecere în revistă pe 146 de pagini a predicțiilor pe care le-am făcut în cărțile mele anterioare, *The Age of Intelligent Machines – Epoca mașinilor inteligente*, *The Age of Spiritual Machines – Epoca mașinilor*

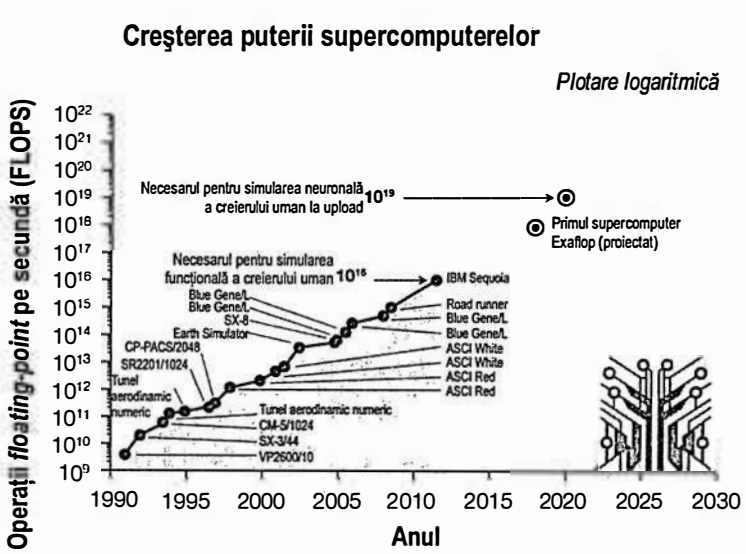
Creșterea exponențială a computației pe 110 ani

Legea lui Moore a fost a cincea, nu prima paradigmă care a adus creștere exponențială în computare

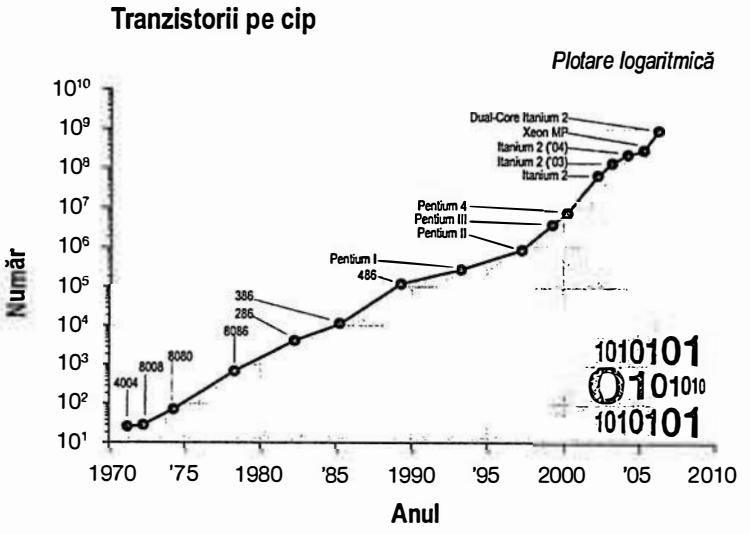
Plotare logaritmică



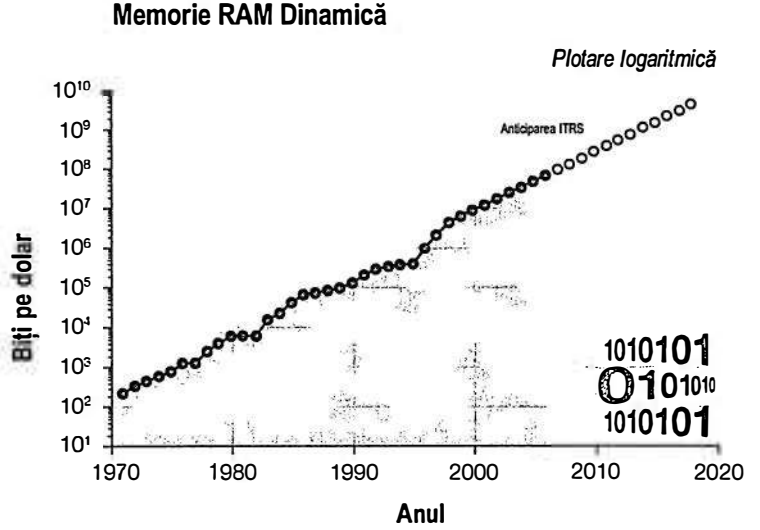
Calcule pe secundă la mia de dolari (constant) ale unor dispozitive de computare[10]



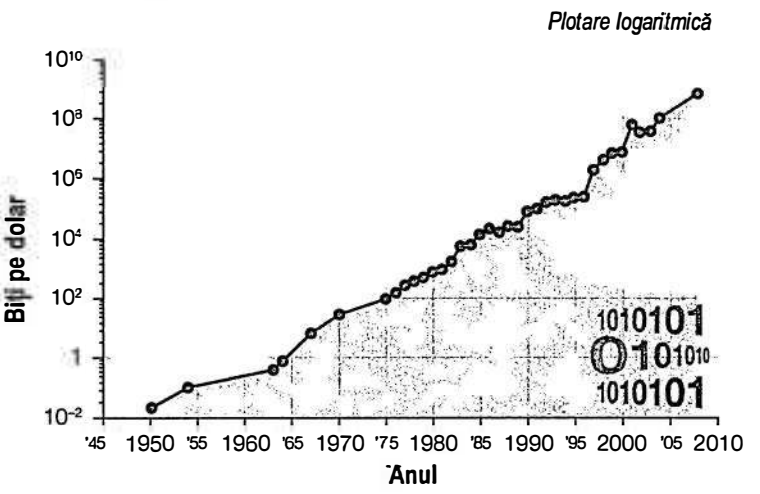
Operaţii *floating-point* pe secundă pe diverse supercomputere[11]



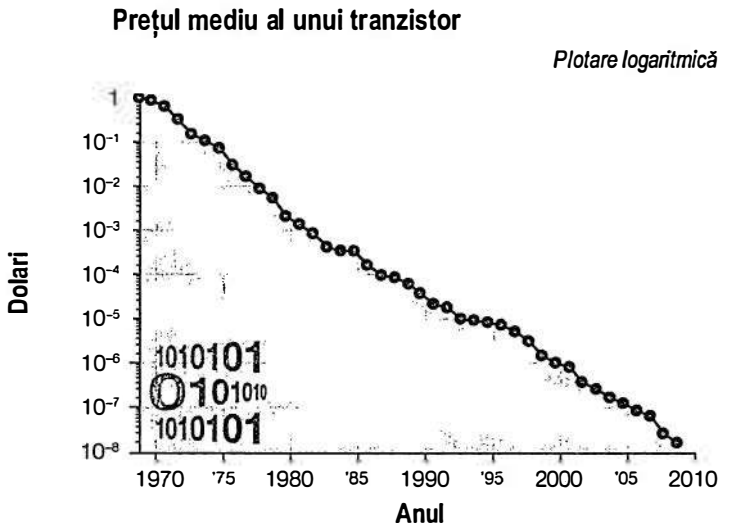
Tranzistori pe cip pentru diverse procesoare Intel[12]



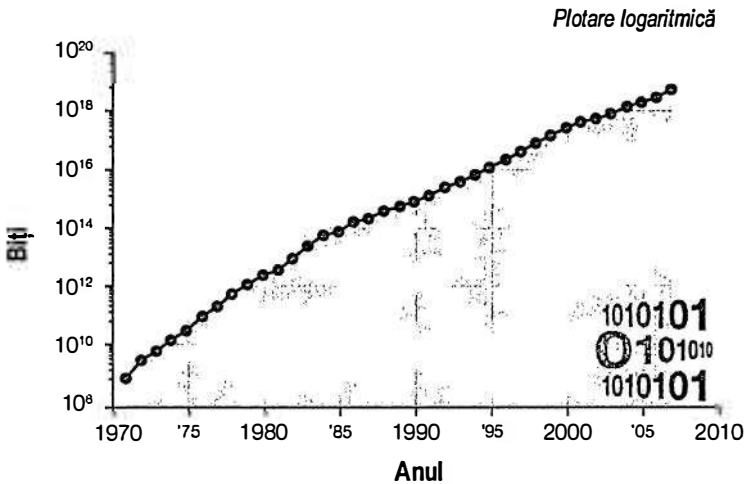
Biți pe dolar pentru cipurile de memorie aleatorie cu acces dinamic[13]



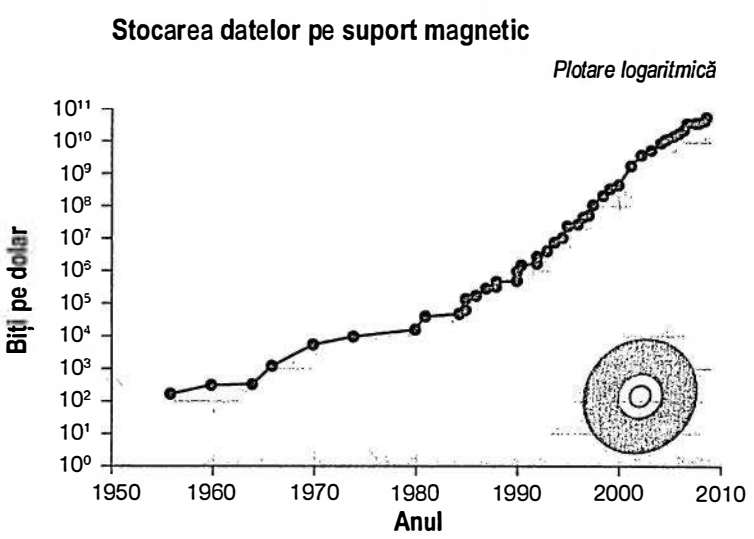
Biți pe dolar pentru cipurile de memorie cu acces aleatoriu[14]



Prețul mediu al unui tranzistor, în dolari[15]



Numărul total de biți de memorie cu acces aleatoriu livrați anual[16]



Biți/dolar (la valoarea constantă de 2.000 USD)
pentru suportul de stocare magnetică a datelor[17]

spirituale şi *The Singularity Is Near – Singularitatea este aproape*. Poți citi eseul aici, mergând la link-ul de la această notă de final.[9] *Epoca maşinilor spirituale* a conținut sute de predicții pentru următoarele decenii: 2009, 2019, 2029 şi 2099. De exemplu, în *The Age of Spiritual Machines – Epoca maşinilor spirituale*, pe care am scris-o în anii 1990, am făcut 147 de predicții pentru anul 2009. Dintre acestea, 115 (78 la sută) sunt absolut corecte la sfârşitul anului 2009; predicțiile bazate pe mărimile capacității şi raportului preț/performanță pentru tehnologiile informației au fost în mod deosebit precise.

Alte 12 predicții (8 la sută) sunt „în esență corecte". Un total de 127 de predicții (86 la sută) sunt corecte ori în esență corecte. (Deoarece predicțiile făcute aveau caracter specific pentru un întreg deceniu, o predicție pentru anul 2009 a fost considerată „în esență corectă" dacă s-a împlinit în 2010 sau 2011.) Încă 17 predicții (12 la sută) sunt parțial corecte şi 3 (2 la sută) sunt greşite.

Nici predicțiile etichetate ca „greşite" nu au fost toate eronate. De exemplu, am considerat greşită predicția mea că o să avem automobile autopilotate, cu toate că Google a făcut demonstrații cu asemenea automobile în luna octombrie a anului 2010, patru camionete electrice fără şofer încheind cu succes un *drive-test* de 13.000 de kilometri din Italia până în China.[18]

Experții din domeniu anticipează actualmente că aceste tehnologii vor fi disponibile în mod curent pentru public înainte de sfârșitul acestui deceniu.

Tehnologiile computației și comunicațiilor, aflate într-o permanentă și exponențială dezvoltare, contribuie, toate, la proiectul de înțelegere și recreare a sistemului creierului uman. Acest efort nu este un singur proiect organizat, ci, dimpotrivă, e rezultatul unei multitudini de proiecte diverse, inclusiv de realizare a modelelor detaliate ale componentelor creierului, mergând de la neuronii individuali la întregul neocortex, la cartografierea „conectomului" (conexiunile neuronale din creier), la simulări are unor regiuni ale creierului și la multe altele. Toate acestea s-au amplificat exponențial. Cea mai mare parte a dovezilor prezentate în această carte au devenit disponibile abia recent – de exemplu, studiul Wedeen din anul 2012, discutat în Capitolul 4, care a dovedit modelul în grilă foarte ordonat și „simplu" (ca să-i citez pe cercetători) pentru conexiunile de la nivelul neocortexului. Cercetătorii care au participat la acest studiu recunosc faptul că intuiția lor (și imaginile) au devenit fezabile ca rezultat al noii tehnologii a imaginilor de înaltă rezoluție.

Tehnologiile de scanare a creierului se îmbunătățesc sub aspectul rezoluției spațiale și temporale într-un ritm exponențial. Diferitele tipuri de scanare a creierului care au fost cercetate merg de la metodele complet neinvazive ce pot fi folosite la om, la metodele mai invazive sau distructive folosite în cazul animalelor.

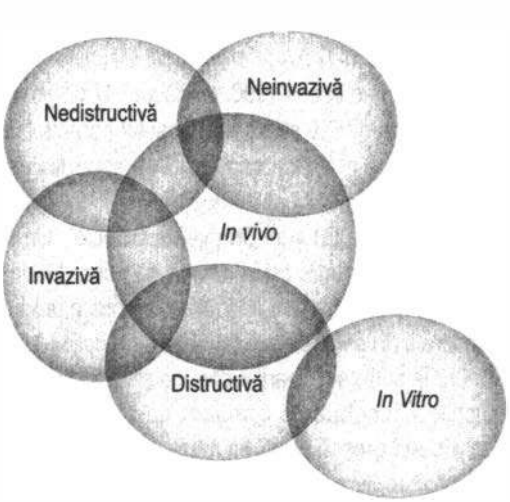
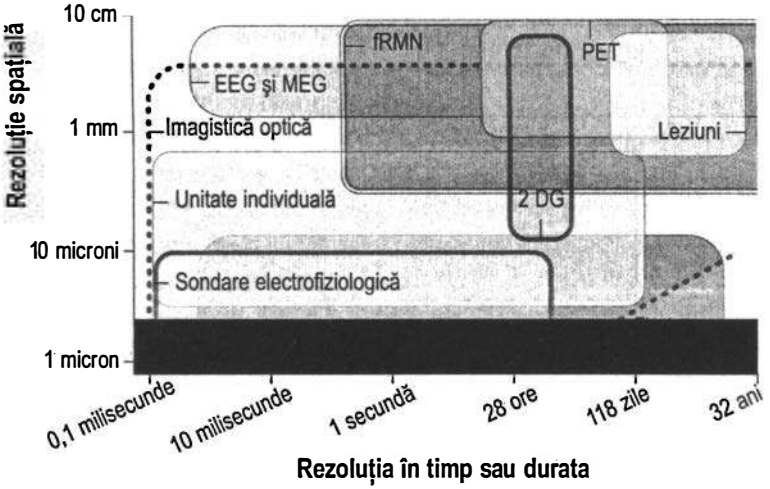
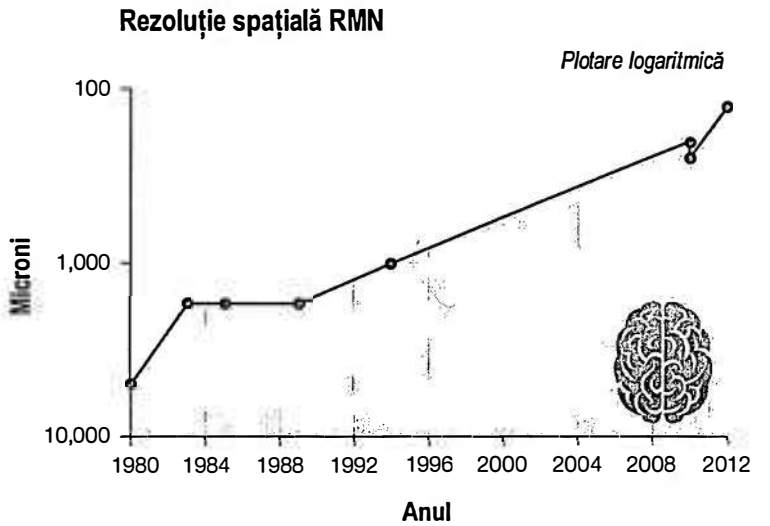


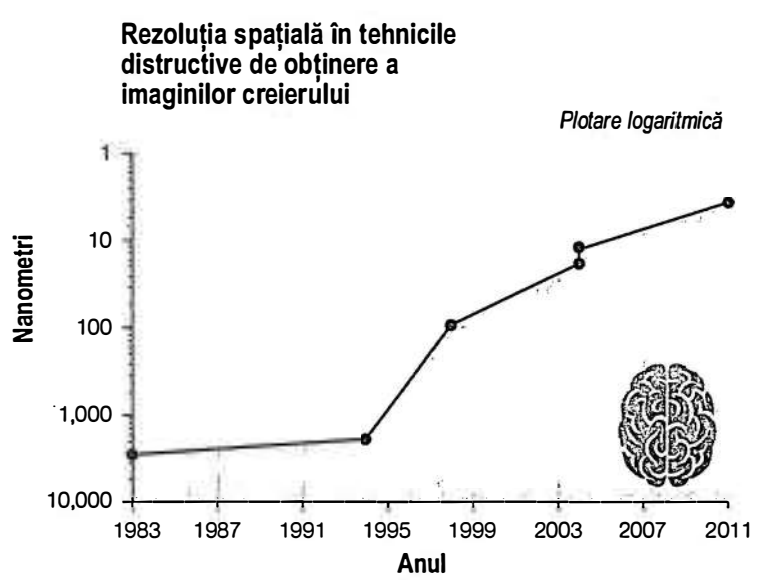
Diagrama Venn privitoare la metodele de obținere a imaginilor creierului[19]



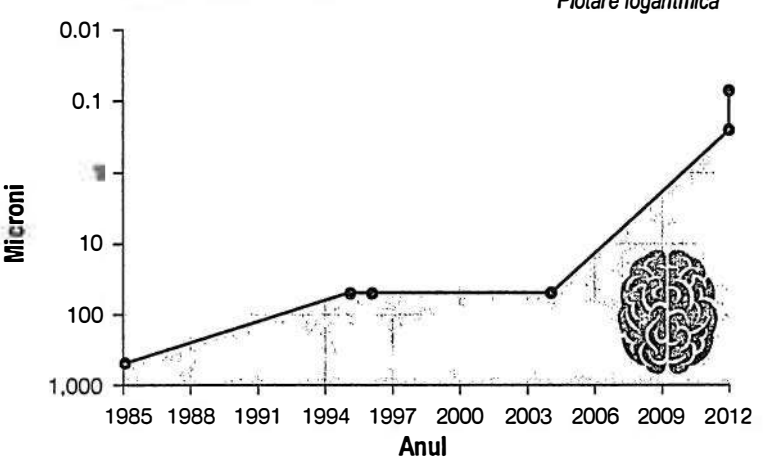
Instrumente pentru obținerea de imagini ale creierului[20]



Rezoluția spațială RMN în microni[21]



Rezoluția spațială la obținerea de imagini ale creierului prin metode distructive[22]



Rezoluția spațială pentru tehnicile nedistructive de obținere a imaginilor creierului la animale[23]

Rezonanța magnetică (RMN), o tehnică imagistică neinvazivă cu o rezoluție relativ înaltă, s-a îmbunătățit permanent, într-un ritm exponențial, până la punctul în care rezoluțiile spațiale sunt acum de aproape 100 de microni (a milioana parte dintr-un metru).

Tehnicile distructive de obținere a imaginilor, care sunt aplicate pentru cartografierea conectomului (harta tuturor conexiunilor interneuronale) din creierul animalelor s-au îmbunătățit şi ele într-un ritm exponențial. Actuala rezoluție maximă este în jur de patru nanometri, adică suficientă pentru a putea fi văzute conexiunile individuale.

Tehnologiile din domeniul inteligenței artificiale, ca de pildă sistemele de înțelegere a limbii naturale, nu sunt concepute neapărat pentru imitarea principiilor teoretice ale funcționării creierului, ci mai curând pentru o eficacitate maximă. În aceste condiții, este notabil faptul că tehnicile care au câştig de cauză sunt consecvente principiilor pe care le-am scos în evidență în această carte: recognizatori autoorganizați, ierarhii de modele invariabile autoasociative cu redundanță şi anticipare ascendentă & descendentă. Şi aceste sisteme se dezvoltă exponențial, după cum a demonstrat Watson.

Un scop principal al înțelegerii creierului este extinderea arsenalului nostru de tehnici pentru crearea sistemelor inteligente. Cu toate că e posibil să nu înțeleagă acest lucru, numeroşi cercetători din domeniul inteligenței artificiale au fost deja influențați profund de cunoştințele pe care le deținem cu privire la principiile de funcționare a creierului. Înțelegerea creierului ne mai ajută şi să îndreptăm disfuncții cerebrale de tot felul. Bineînțeles, acesta este un alt scop-cheie al proiectului de abordare a creierului prin metodele retroingineriei: să înțelegem cine suntem.

Capitolul 11

OBIECȚII

> Dacă o mașină s-ar putea dovedi imposibil de deosebit de un om, ar trebui să-i acordăm respectul pe care l-am arăta unui om – ar trebui să acceptăm că are minte.
>
> **— Stevan Harnad**

Cea mai semnificativă sursă de obiecții la teza mea reprezentată de Legea accelerării veniturilor și de aplicarea ei în domeniul amplificării inteligenței umane își are originile în natura liniară a intuiției omului. După cum voi arăta în continuare, fiecare dintre cele câteva sute de milioane de recognizatori de modele de la nivelul neocortexului procesează informația secvențial. Una dintre implicațiile acestei organizări este aceea că, în privința viitorului, avem așteptări liniare, prin urmare criticii își aplică intuiția liniară la fenomenele informației, care sunt eminamente exponențiale.

Numesc obiecțiile de acest tip „critici din lipsă de încredere", datorită faptului că proiecțiile exponențiale par incredibile ținând cont de predilecția noastră liniară și îmbracă o varietate de forme. Cofondatorul Microsoft Paul Allen (născut în anul 1953) și colegul său Mark Greaves au formulat recent o serie de obiecții de acest gen, într-un eseu intitulat *The Singularity Isn't Near – Singularitatea nu este aproape*, publicat în revista *Technology Review*.[1] Răspunsul meu de aici este destinat doar criticilor lui Allen, dar ele reprezintă o gamă tipică de obiecții la argumentele pe care le-am avansat în special în legătură cu creierul. În ciuda referirii lui Allen, în titlul eseului, la cartea mea *The Singularity Is Near – Singularitatea este aproape*, singurul lui citat provine dintr-un eseu pe care l-am scris în anul 2001 (intitulat *The Law of Accelerating Returns – Legea accelerării veniturilor*). Mai mult chiar, articolul lui nu recunoaște și nici nu ține seamă de argumentele pe care le-am prezentat în carte. Din păcate, mult prea des constat că așa stau lucrurile în cazul criticilor aduse cercetărilor mele.

Când a fost publicată, în anul 1999, fiind îmbogățită ulterior cu eseul din anul 2001, cartea mea *The Age of Spiritual Machines – Epoca mașinilor spirituale* a generat critici pe mai multe direcții, cum ar fi acelea că: Legea

lui Moore va ajunge la sfârşit; capacitatea hardware se va perfecționa exponențial, dar partea software e împotmolită; creierul este mult prea complicat; există capacități ale creierului care, în mod inerent, nu pot fi reproduse de un software; şi alte câteva. Unul dintre motivele pentru care am scris cartea *The Singularity Is Near* a fost să răspund acestor critici.

Nu pot spune că Allen şi alți critici asemenea lui ar fi trebuit neapărat să se lase convinşi de argumentele pe care le-am prezentat în această carte, dar, cel puțin, şi el, şi ceilalți ar fi putut răspunde la ceea ce am scris eu efectiv. Allen susține că „Legea accelerării veniturilor... nu este o lege fizică". Aş scoate în evidență faptul că partea cea mai însemnată a legilor din știință nu sunt legi fizice; ele rezultă din proprietățile emergente ale unui mare număr de evenimente petrecute la nivel mai redus. Un exemplu clasic este reprezentat de Legile termodinamicii. Dacă te uiți pe suportul matematic aflat la baza Legilor termodinamicii, observi că ele realizează pentru fiecare particulă un model care urmează o cale aleatorie, astfel încât, prin definiție, nu putem anticipa unde anume se va afla o anumită particulă într-un moment din viitor. Cu toate acestea, proprietățile de ansamblu ale gazului sunt cât se poate de predictibile, cu un grad înalt de precizie, potrivit Legilor termodinamicii. La fel se întâmplă şi cu Legea accelerării veniturilor: orice proiect şi contributor din domeniul tehnologic e imprevizibil, însă traiectoria de ansamblu, aşa cum este ea cuantificată de mărimea raportului preț/performanță şi capacitate, urmează, totuşi, o cale remarcabil de predictibilă.

Dacă tehnologia computerelor ar reprezenta preocuparea unui grup restrâns de cercetători, ea ar fi într-adevăr imprevizibilă. Însă este produsul unui sistem de proiecte suficient de dinamic pentru ca raportul fundamental preț/performanță (de exemplu calcule pe secundă per dolar constant) să urmeze o cale exponențială foarte netedă, datând încă de la recensământul american din anul 1890, aşa cum am subliniat în capitolul anterior. Deşi baza teoretică pentru Legea accelerării veniturilor este prezentată pe larg în cartea mea *The Singularity Is Near – Singularitatea este aproape*, demonstrațiile cele mai solide în sprijinul ei vin din dovezile empirice pe care le-am prezentat eu şi alții.

Allen scrie că „aceste «legi» merg până nu mai merg". Aici, el confundă paradigmele cu traiectoria continuă a unui domeniu de bază al tehnologiei informației. Dacă am examina, de exemplu, tendința construirii unor tuburi cu vid din ce în ce mai mici – paradigma progresului computației în

anii 1950 –, am spune că e adevărat că ea a mers până nu a mai mers. Dar, atunci când finalul acestei paradigme a devenit evident, a crescut presiunea la care a fost supusă cercetarea pentru emiterea următoarei paradigme. Tehnologia tranzistorilor a păstrat tendința de creştere exponențială a raportului preț/performanță în domeniul computației şi asta a condus la a cincea paradigmă (Legea lui Moore) şi la continua comprimare a particularităților pe circuite integrate. S-au emis permanent predicții referitoare la faptul că Legea lui Moore va ajunge la un final. Setul de documente intitulat *International Technology Roadmap for Semiconductors – Ghidul tehnologic internațional pentru semiconductori* proiectează particularități de şapte nanometri până la începutul anilor 2020.[2] În acest punct, particularitățile-cheie vor avea diametrul de treizeci şi cinci de atomi de carbon şi va fi dificil să fie comprimate şi mai mult. Pe de altă parte însă, Intel şi alți producători de cipuri întreprind deja primii paşi în direcția celei de-a şasea paradigme, computarea în trei dimensiuni, pentru continuarea îmbunătățirii exponențiale a raportului preț/performanță. Intel intenționează să transforme cipurile tridimensionale în ceva obişnuit într-un orizont de timp de 13-19 ani; deja au fost lansați tranzistorii tridimensionali şi cipurile de memorie 3D. Această a şasea paradigmă va continua să crediteze Legea accelerării veniturilor în privința raportului preț/performanță până mai târziu în decursul acestui secol, când un volum de computație în valoare de o mie de dolari va fi de miliarde de ori mai puternic decât creierul omului.[3] (Se pare că Allen şi cu mine ne punem în sfârşit de acord în privința nivelului de computație necesar pentru simularea funcțională a creierului uman.)[4]

Allen merge apoi mai departe şi avansează argumentul-standard că software-ul nu progresează în aceeaşi manieră exponențială ca hardware-ul. În cartea mea *The Singularity Is Near – Singularitatea este aproape*, m-am ocupat pe larg de acest subiect, citând diverse metode de măsurare a complexității şi capacității programelor software, care într-adevăr demonstrează o creştere exponențială similară.[5] Un studiu recent (*Report to the President and Congress, Designing a Digital Future: Federally Funded Research and Development in Networking and Information Technology – Raport pentru preşedinte şi Congres privind conceperea unui viitor digital: cercetare şi dezvoltare finanțată de la nivel federal pentru networking şi tehnologia informației* – aparținând grupului de consilieri ai preşedintelui pentru ştiință şi tehnologie) afirmă următoarele:

Şi mai remarcabil – şi chiar mai puțin înțeles – este faptul că, în numeroase domenii, performanța are câștig de cauză pentru că perfecționarea algoritmilor a depăşit cu mult chiar şi performanțele spectaculoase datorate creşterii vitezei procesoarelor. Algoritmii pe care îi folosim astăzi pentru recunoaşterea vorbirii, pentru traducerea în limbă naturală, pentru jocul de şah sau pentru planificarea logistică au evoluat remarcabil în ultimul deceniu... Iată un singur exemplu, oferit de profesorul Martin Grötschel de la Konrad-Zuse-Zentrum für Informationstechnik din Berlin. Grötschel, expert în optimizare, observă că realizarea unui model de planificare pentru o linie de producție cu ajutorul programării liniare ar fi avut nevoie de 82 de ani în 1988, folosindu-se computerele şi algoritmii de programare liniară ai vremii. Cincisprezece ani mai târziu – în anul 2003 –, acelaşi model putea fi realizat în doar circa un minut, ceea ce înseamnă o îmbunătățire de aproximativ 43 de milioane de ori. Din această sumă, o creştere de aproximativ 1.000 de ori s-a datorat măririi vitezei procesorului, iar restul de 43.000 de ori a fost datorată progreselor privitoare la algoritmi! Grötschel mai citează şi o îmbunătățire de circa 30.000 de ori a algoritmilor pentru programarea mixtă a întregilor între anii 1991 şi 2008. Proiectarea şi analiza algoritmilor, ca şi studiul complexității computaționale inerente a problemelor sunt subdomenii fundamentale ale științei computerelor.

Reține, cititorule, că programarea liniară, citată mai înainte de Grötschel ca fiind beneficiara unei îmbunătățiri a performanțelor de 43 de milioane la 1, este tehnica matematică folosită pentru atribuirea optimă a resurselor într-un sistem de memorie ierarhică de felul Modelelor Markov Ascunse Ierarhic despre care am vorbit anterior. Şi citez şi alte asemenea exemple în cartea mea *The Singularity Is Near*.[6]

În privința inteligenței artificiale, Allen se grăbeşte să-l respingă pe Watson, creația IBM – opinie împărtăşită de numeroşi alți critici. Mulți dintre aceşti detractori nu ştiu absolut nimic despre Watson în afara faptului că e un software care rulează pe un computer (deşi este unul paralel cu 720 de procesoare). Allen scrie că sisteme precum Watson „rămân fragile, limitele performanțelor lor sunt stabilite rigid de premisele şi algoritmii lor definitorii interni şi adeseori dau răspunsuri fără sens în afara domeniilor lor specifice".

În primul rând, am putea face o observație similară in privința omului. De asemenea, aş sublinia că „domeniile specifice" ale lui Watson includ

toată Wikipedia, plus multe alte baze de cunoştinţe care cu greu ar putea fi considerate un obiectiv îngust. Watson se ocupă de o gamă bogată de cunoştinţe umane şi este capabil să facă faţă unor forme subtile ale limbajului, inclusiv jocuri de cuvinte, comparaţii şi metafore, practic pentru orice domeniu al strădaniilor omeneşti. Nu este perfect, dar nici oamenii nu sunt, iar el a fost suficient de bun ca să câştige la *Jeopardy!* în faţa celor mai buni jucători umani.

Allen susţine că Watson a fost asamblat personal de către oamenii de ştiinţă, care au construit fiece legătură de cunoaştere îngustă în domenii specifice. Pur şi simplu, acest fapt nu este adevărat. Deşi câteva domenii ale datelor lui Watson au fost programate direct, Watson şi-a dobândit singur majoritatea semnificativă a cunoştinţelor citind în limbaj natural documente cum ar fi Wikipedia. Acesta este punctul-forte al lui Watson, iar lui i se adaugă capacitatea de a înţelege limbajul ocolit al întrebărilor de la *Jeopardy!* (răspunsuri în căutarea unei întrebări).

După cum am menţionat anterior, în mare, criticile la adresa lui Watson vizează faptul că acesta funcţionează pe baza probabilităţilor statistice şi nu a unei „adevărate" înţelegeri. Mulţi cititori pricep că asta înseamnă că Watson nu face decât să adune statistici privitoare la succesiunile cuvintelor. Sintagma „informaţie statistică", în cazul lui Watson, se referă de fapt la coeficienţii distribuiţi şi la conexiunile simbolice ale metodelor autoorganizate, cum ar fi Modelele Markov Ascunse Ierarhice. La fel de uşor se poate respinge ideea potrivit căreia concentraţiile neurotransmiţătorilor şi modelele de conexiuni redundante de la nivelul cortexului uman sunt „informaţii statistice". Da, şi noi rezolvăm ambiguităţile într-un mod foarte asemănător cu Watson – apreciind probabilitatea diverselor interpretări ale unei expresii.

Allen continuă: „Orice structură [din creier] a fost precis adecvată de milioanele de ani de evoluţie pentru a face un anumit lucru, indiferent care este acesta. Nu e ca un computer, cu milioane de tranzistori identici în dispoziţii regulate de memorie controlate de o unitate centrală de procesare cu câteva elemente diferite. La nivelul creierului, toate structurile individuale şi toate circuitele neuronale au fost rafinate individual de evoluţie şi de factorii de mediu".

Această afirmaţie că toate structurile şi toate circuitele neuronale ale creierului sunt unice şi se află intenţionat acolo este pur şi simplu imposibilă, căci ar însemna că proiectul creierului ar impune sute de mii de miliarde de biţi de informaţie. Planul structural al creierului (asemenea celui al restului

corpului) este conținut în genom și creierul însuși nu poate să depoziteze mai multă informație de concepție decât genomul. Observă, cititorule, că informația epigenetică (de pildă controlul expresiei genelor de către peptide) nu contribuie în mod apreciabil la volumul de informație de la nivelul genomului. Experiența și învățarea sporesc într-adevăr semnificativ volumul de informație conținut de creier, dar se pot spune și despre inteligența artificială aceleași lucruri care se spun despre Watson. În cartea mea *The Singularity is Near – Singularitatea este aproape*, arăt că, după comprimarea fără pierderi (datorată redundanței masive a genomului), volumul de informație de concepție din genom se ridică la aproximativ 50 de milioane de biți, din care în jur de jumătate (respectiv circa 25 de milioane de biți) are legătură cu creierul.[7] Nu este ceva simplu, însă e un nivel de complexitate căruia îi putem face față și reprezintă o complexitate mai redusă decât a multor sisteme software din lumea modernă. În plus, peste 25 de milioane de biți din informația generică de concepție a creierului au legătură cu cerințele biologice pentru neuroni, nu cu algoritmii lor de funcționare.

Cum ajungem la un număr de 100 până la 1.000 de mii de miliarde de conexiuni la nivelul creierului de pe urma câtorva zeci de milioane de biți de informație de concepție? În mod evident, răspunsul este: prin redundanță masivă. Dharmendra Modha, managerul de Computație Cognitivă de la IBM Research, scrie că „neuroanatomiștii nu au descoperit o rețea incredibil de încâlcită, arbitrar conectată, cu totul deosebită, în creierul fiecărui individ, ci, dimpotrivă, foarte multe structuri care se repetă și o mare analogie între specii... Reconfigurabilitatea naturală uluitoare dă speranțe că algoritmii de bază pentru neurocomputație sunt independenți de modalitățile senzoriale și motorii și că mare parte din variațiile observate la nivelul structurii corticale a tuturor regiunilor reprezintă o rafinare a unui circuit canonic; și tocmai acest circuit canonic este cel pe care vrem să-l copiem prin retroinginerie".[8]

Allen pledează în favoarea unei inerente „frâne de complexitate, care în mod necesar ar limita progresul înțelegerii creierului uman și replicarea capacităților lui", pe baza ideii că toate cele aproximativ 100 până la 1.000 de mii de miliarde de conexiuni din creierul uman există datorită unui proiect explicit. Expresia sa „frână de complexitate" confundă pădurea cu copacii. Dacă vrei să înțelegi, să realizezi modele, să simulezi și să recreezi un pancreas, nu trebuie să refaci sau să simulezi fiecare organită din fiecare celulă a insulelor pancreatice Langerhans. Ai vrea, în schimb, să înțelegi o

asemenea celulă, să-i abstractizezi funcționalitatea fundamentală legată de controlul nivelului insulinei și pe urmă să o extinzi la un grup numeros de asemenea celule. Acest algoritm este bine înțeles în ceea ce privește celulele din insulele Langerhans. Actualmente există pancreasuri artificiale ce utilizează acest model funcțional care este testat. Deși cu siguranță există o complexitate și o variabilitate cu mult mai mari la nivelul creierului comparativ cu complexitatea și variabilitatea celulelor din insulele Langerhans ale pancreasului, se constată pe de altă parte și o masivă repetare a funcțiilor, așa cum am arătat adesea în această carte.

Criticii care merg pe linia lui Allen mai manifestă și ceea ce eu aș numi „pesimismul omului de știință". Cercetătorii implicați în lucrul la următoarea generație a unei tehnologii sau a modelelor dintr-un domeniu științific se luptă în mod invariabil cu acel set imediat de provocări, astfel încât, dacă cineva descrie felul cum va arăta tehnologia peste zece generații, le sclipesc ochii. Unul dintre pionierii circuitelor integrate îmi reamintea recent de strădaniile de trecere de la mărimi ale particularităților de 10 microni (10.000 de nanometri) la mărimi de 5 microni (5.000 de nanometri), în urmă cu mai bine de treizeci de ani. Oamenii de știință erau încrezători cu prudență în atingerea acestui scop, dar, când au anticipat că într-o bună zi vom avea într-adevăr circuite cu dimensiunile particularităților sub un micron (1.000 de nanometri), cei mai mulți dintre ei, concentrați asupra scopului, s-au gândit că era un lucru prea exagerat ca să se oprească asupra lui. Au fost ridicate obiecții referitoare la fragilitatea circuitelor la acel nivel de precizie, la efectele termice și așa mai departe. Astăzi, Intel începe să folosească cipuri cu lungimi de poartă de 22 de nanometri.

Am fost martorii aceluiași gen de pesimism în privința Proiectului Genomului Uman. La jumătatea efortului cu durata de cincisprezece ani, doar 1 la sută din genom fusese achiziționat, iar criticii propuneau limite elementare privind viteza cu care putea fi secvențializat acesta fără distrugerea delicatelor structuri genetice. Însă, mulțumită creșterii exponențiale a capacității și a raportului preț/performanță, proiectul a fost încheiat șapte ani mai târziu. Proiectul de copiere a creierului uman prin retroinginerie progresează în mod similar. Abia recent, de exemplu, am ajuns cu tehnicile neinvazive la un prag care să ne permită să observăm în timp real conexiunile interneuronale individuale atunci când se formează și generează impulsurile nervoase. Cele mai multe dintre dovezile pe care le-am prezentat în cartea de față au depins de aceste evoluții și abia recent au devenit disponibile.

Allen mi-a descris propunerea de recreare a creierului uman prin retroinginerie ca pe o simplă scanare a creierului pentru înțelegerea structurii sale fine, urmată de simularea întregului creier „pe invers", fără înțelegerea metodelor lui de procesare a informațiilor. Trebuie neapărat să înțelegem în detaliu cum funcționează diferitele tipuri individuale de neuroni și abia apoi să adunăm informația cu privire la conectarea modulelor funcționale. Metodele derivate din acest gen de analiză pot apoi să ghideze dezvoltarea unor sisteme inteligente. În esență, căutăm metode inspirate de biologie care pot accelera munca în domeniul inteligenței artificiale. Această muncă a progresat în mare parte în absența unor intuiții semnificative privitoare la modul în care creierul îndeplinește funcții similare. Din cercetările mele personale în domeniul recunoașterii vorbirii, știu că munca ne-a fost accelerată foarte mult din momentul când am beneficiat de o intuiție referitoare la modul în care creierul pregătește și transformă informația auditivă.

Modul în care se diferențiază structurile masiv redundante din creier rezidă în învățare și experiență. Vârful actual al inteligenței artificiale le permite, de fapt, sistemelor să învețe și din propria experiență. Automobilele Google care se pilotează singure învață din propria lor experiență de pilotaj și din datele automobilelor Google pilotate de șoferi umani; Watson și-a acumulat cea mai mare parte a cunoștințelor citind singur. Este interesant de observat că metodele folosite astăzi în domeniul inteligenței artificiale au evoluat pentru a fi matematic foarte similare cu mecanismele de la nivelul neocortexului.

O altă obiecție la adresa fezabilității „inteligenței artificiale puternice" (inteligență artificială la nivel uman și peste) susține că creierul uman folosește extensiv computarea analogică, în vreme ce metodele digitale nu pot replica în mod inerent gradațiile de valoare pe care le pot întruchipa reprezentările analogice. Este adevărat că un bit este fie deschis, fie închis, însă cuvintele compuse din mai mulți biți reprezintă cu ușurință mai multe gradații și o pot face cu orice grad de acuratețe dorit. Desigur, computerele digitale fac acest lucru permanent. Ca urmare, acuratețea informației analogice la nivelul creierului (tăria sinaptică, de exemplu) este doar de un nivel dintre cele 256 de niveluri de opt biți care pot fi reprezentate.

În Capitolul 9, am citat obiecția lui Roger Penrose și Stuart Hameroff referitoare la microtubuli și la computarea cuantică. Îți amintești, cititorule: ei afirmau că structurile de microtubuli din neuroni realizează o computare cuantică, așa încât, de vreme ce nu este posibil să faci asta în cazul

computerelor, creierul uman e fundamental diferit şi, după câte s-ar putea presupune, mai bun. Aşa cum am argumentat mai înainte, nu există nicio dovadă că microtubulii neuronali realizează computarea cuantică. De fapt, omul se descurcă destul de slab la rezolvarea genului de probleme la care un computer cuantic ar excela (cum ar fi înmulţirea unor numere mari). Şi, dacă s-ar dovedi a fi adevărat ceva din toate acestea, nimic nu ar putea să împiedice folosirea computării cuantice şi în cazul computerelor.

John Searle este celebru pentru lansarea experimentului mintal pe care el îl numeşte „camera chinezească", un argument pe care îl discut detaliat în cartea mea *The Singularity is Near*.[9] Pe scurt, este vorba despre un bărbat care primeşte întrebări scrise în limba chineză şi apoi răspunde la ele. Pentru asta, foloseşte o carte complicată de reguli. Searle susţine că bărbatul nu înţelege cu adevărat limba chineză şi că nu este „conştient" în limba respectivă (deoarece nu înţelege întrebările şi răspunsurile), în ciuda aparentei sale abilităţi de a răspunde la întrebări. Searle compară această situaţie cu cea a unui un computer şi concluzionează că un computer care ar putea răspunde la întrebări puse în chineză (în esenţă trecând un test Turing în limba chineză), asemenea bărbatului din camera chinezească, nu ar înţelege în mod real limba şi nici nu ar fi conştient de ceea ce face.

Există mai multe scăpări filosofice în argumentaţia lui Searle. În primul rând, bărbatul din acest experiment mintal este comparabil doar cu unitatea centrală de procesare a unui computer. S-ar putea spune că unitatea centrală de procesare nu înţelege cu adevărat ce face, însă ea e doar o parte a structurii. În camera chinezească a lui Searle, bărbatul cu cartea de reguli în mână constituie întregul sistem. Acest sistem înţelege limba chineză; altminteri, nu ar fi capabil să răspundă convingător la întrebările puse în această limbă, fapt ce ar viola ipoteza lui Searle din experimentul său mintal.

Atractivitatea argumentului lui Searle rezidă în faptul că, în zilele noastre, este dificil să se deducă ideea că un software ar înţelege cu adevărat sau ar avea conştienţă. Însă problema acestui argument este că poţi să aplici raţionamentul său şi creierului uman. Toţi recognizatorii neocorticali de modele – şi chiar toţi neuronii şi toate componentele fiecărui neuron – respectă un algoritm. (La urma urmelor, sunt mecanisme moleculare, care respectă legea naturală.) Dacă am concluziona că respectarea unui algoritm este incompatibilă cu adevărata înţelegere şi conştienţă, atunci ar trebui să concluzionăm că nici creierul uman nu prezintă aceste calităţi. Poţi lua argumentul lui Searle cu camera chinezească şi pur şi simplu poţi înlocui

sintagma lui „manipularea simbolurilor" cu „manipularea conexiunilor interneuronale şi a tăriei sinaptice". Vei obține un argument cu adevărat convingător în sensul că creierul uman realmente nu poate să înțeleagă nimic.

O altă direcție a argumentării vine din însăşi natura naturii, care a devenit un nou teren sacru pentru mulți observatori. De exemplu, biologul neo-zeelandez Michael Denton (născut în anul 1943) vede o diferență profundă între principiile de concepere a maşinilor şi cele ale biologiei. Denton scrie că entitățile naturale sunt „autoorganizate, ...autoreferențiale, ...autoreplicative, ...autoformatoare şi... holistice".[10] El susține că asemenea forme biologice nu pot fi create decât prin procese biologice şi că, prin urmare, ele sunt realități „imuabile, ...impenetrabile şi... fundamentale" ale existenței, ceea ce înseamnă că reprezintă o categorie filosofică diferită de a maşinilor.

Realitatea, aşa cum am văzut, este că pot fi concepute şi maşini pe baza aceloraşi principii. A învăța paradigmele specifice de concepție ale celei mai inteligente entități a naturii – creierul uman – este exact scopul proiectului de realizare a creierului prin retroinginerie. De asemenea, e neadevărat faptul că sistemele biologice sunt complet „holistice", după cum se exprimă Denton şi nici că, reciproc, maşinile trebuie să fie complet modulare. Am identificat în mod clar ierarhii ale unităților de funcționalitate în sistemele naturale, în special în creier, iar sistemele de inteligență artificială folosesc metode comparabile.

Îmi face impresia că mulți critici vor rămâne nemulțumiți până când computerele vor trece Testul Turing în mod curent, însă nici acest prag nu va fi clar. Fără îndoială, vor apărea controverse legate de faptul că pretinsele Teste Turing administrate ar fi valide. Fără îndoială, mă voi număra probabil printre acei critici care vor discredita afirmațiile timpurii pe aceste direcții. Până se vor linişti disputele referitoare la validitatea unui computer care trece Testul Turing, computerele vor fi depăşit de mult inteligența umană neasistată.

Accentul meu aici este pus pe cuvântul „neasistată", deoarece asistarea este exact motivul pentru care creăm aceşti „copii ai minții", cum îi numeşte Hans Moravec.[11] Combinarea recogniției de modele de nivel uman cu viteza inerentă şi acuratețea computerelor se va concretiza prin abilități redutabile. Însă nu este vorba despre o invazie extraterestră a maşinilor inteligente de pe Marte – creăm aceste instrumente pentru a deveni noi înşine mai inteligenți. Cred că majoritatea observatorilor vor fi de acord cu mine că acesta este aspectul unic al speciei umane: construim aceste instrumente pentru a ne extinde propria anvergură.

EPILOGUL

Imaginea e foarte anostă, domnilor... În lume clima e în schimbare, mamiferele preiau supremația şi avem cu toții un creier cât o nucă.

— *Dinozaurii care vorbesc*, în *The Far Side* de Gary Larson

Inteligența poate fi definită drept abilitatea de a rezolva probleme cu ajutorul unor resurse limitate, dintre care o resursă-cheie o reprezintă timpul. Prin urmare, abilitatea de a rezolva mai rapid o problemă cum ar fi găsirea hranei sau evitarea unui prădător reflectă o forță mai mare a intelectului. Inteligența a evoluat deoarece a fost utilă pentru supraviețuire – fapt ce poate părea evident, dar cu care nimeni nu este de acord. Aşa cum a fost aplicată de specia noastră, inteligența ne-a permis nu numai să dominăm planeta, ci şi să ne îmbunătățim permanent calitatea vieții. Nici acest din urmă aspect nu este evident pentru toată lumea, dat fiind faptul că astăzi există percepția larg răspândiră că viața devine din ce în ce mai grea. De exemplu, un sondaj Gallup dat publicității în data de 4 mai 2011 a dezvăluit că numai „44 la sută dintre americani cred că tinerii de astăzi vor avea o viață mai bună decât părinții lor".[1]

Dacă ne uităm la tendințele mai ample, nu numai că speranța de viață a omului s-a mărit de patru ori pe parcursul ultimului mileniu (şi a crescut de mai mult de două ori în cursul ultimelor două secole)[2], dar şi PIB-ul pe cap de locuitor (în dolari constanți la valoarea curentă) a crescut de la câteva sute de dolari în anul 1800, la mii de dolari în zilele noastre, cu tendințe şi mai pronunțate în lumea dezvoltată.[3] În urmă cu un secol, existau doar câteva democrații, în vreme ce azi ele reprezintă standardul. Pentru a căpăta o perspectivă istorică privitoare la proporțiile progresului nostru, îți sugerez să citeşti *Leviatanul* lui Hobbes (1651), în care autorul descrie „viața omului" ca fiind „solitară, săracă, grea, brutală şi scurtă". Pentru o perspectivă modernă, recenta carte *Abundance* (2012) a întemeietorului X-Prize Foundation (şi cofondator alături de mine al Singularity University) Peter Diamandis şi a autorului pe teme ştiințifice Steven Kotler documentează modurile extraordinare în care viața din zilele noastre s-a îmbunătățit permanent pe toate

dimensiunile ei. Recenta carte a lui Steven Pinker *The Better Angels of Our Nature: Why Violence Has Declined* (2011) probează minuțios amplificarea relațiilor pașnice dintre oameni și popoare. Avocatul, antreprenorul și autorul american Martine Rothblatt (născută în anul 1954) documentează îmbunătățirea permanentă a drepturilor civile, subliniind, de exemplu, că, în câteva decenii, căsătoriile homosexuale au ajuns de la stadiul de a nu fi acceptate niciunde în lume la recunoașterea legală într-un număr tot mai mare de jurisdicții.[4]

Un motiv principal pentru care oamenii cred că viața din zilele noastre este din ce în ce mai grea rezidă în faptul că informațiile privitoare la problemele lumii s-au amplificat permanent. În zilele noastre, dacă are loc o bătălie undeva pe Pământ, trăim experiența ei aproape ca și cum ne-am afla acolo. În timpul celui de-al Doilea Război Mondial, puteau să piară zeci de mii de oameni într-o bătălie și, în cazul în care publicul ajungea să afle, acest lucru se întâmpla odată cu vizionarea unui jurnal plin de purici, într-un cinematograf, la distanță de câteva săptămâni. În timpul Primului Război Mondial, o elită redusă numeric putea să citească despre evoluția conflictului în ziar (fără imagini). În secolul al nouăsprezecelea, aproape că nu exista acces la știri de actualitate.

Progresul pe care l-am înregistrat ca specie datorită inteligenței noastre este reflectat în evoluția cunoașterii, care include tehnologia și cultura noastră. Diversele tehnologii ale omului devin într-o măsură din ce în ce mai mare tehnologii ale informației, continuând inerent progresul într-o manieră exponențială. Prin asemenea tehnologii, putem răspunde marilor provocări ale omenirii, cum ar fi păstrarea unui mediu sănătos, asigurarea resurselor pentru creșterea populației (inclusiv energie, alimente și apă), prevenirea și vindecarea bolilor, creșterea amplă a longevității omului și eradicarea sărăciei. Numai extinzându-ne cu tehnologii inteligente putem atinge complexitatea necesară pentru înfruntarea acestor provocări.

Aceste tehnologii nu sunt avangarda unei invazii inteligente care va concura cu noi și în cele din urmă ne va disloca. De când am pus mâna pe un băț ca să ajungem la o creangă mai de sus, ne-am folosit uneltele pentru a ne extinde anvergura, atât fizic, cât și mental. Faptul că astăzi putem scoate din buzunar un dispozitiv și putem accesa cunoștințele umane cu câteva tastări ne duce dincolo de tot ce era imaginabil pentru majoritatea observatorilor cu numai câteva decenii în urmă. „Telefonul celular" (termenul este pus în ghilimele pentru că e vorba de ceva cu mult mai mult decât de un

simplu telefon) din buzunarul meu este de un milion de ori mai ieftin, dar de o mie de ori mai puternic decât computerul pe care îl împărțeau toți profesorii și studenții de la MIT pe când studiam acolo. Asta înseamnă o creștere de un miliard de ori a raportului preț/performanță în decursul ultimilor patruzeci de ani, ascensiune la care vom fi din nou martori în următorii douăzeci și cinci de ani, când ceea ce încăpea într-o clădire și acum ne încape în buzunar va încăpea într-un element figurat sanguin.

În acest fel, vom fuziona cu tehnologia inteligentă pe care o creăm. Nanoboții inteligenți din circulația noastră sanguină ne vor menține corpurile biologice sănătoase la nivel celular și molecular. Vor pătrunde în creier în mod neinvaziv, prin capilare și vor interacționa cu neuronii noștri biologici, amplificându-ne în mod direct inteligența. Nu este ceva atât de futurist pe cât ar putea părea. Există numeroase dispozitive de mărimea unui element figurat al sângelui care pot să vindece diabetul de tipul I la animale sau să detecteze și să distrugă celulele canceroase din circulația sanguină. Pe baza Legii accelerării veniturilor, aceste tehnologii vor fi de un miliard de ori mai eficiente în următoarele trei decenii decât sunt acum.

Realmente consider dispozitivele pe care le folosesc și norul de resurse de computare la care sunt ele conectate virtual ca fiind extensii ale mele și mă simt incomplet dacă mă văd despărțit de aceste prelungiri ale creierului. Iată motivul pentru care greva de o zi de la Google, Wikipedia și de la mii de alte website-uri împotriva SOPA (Stop Online Piracy Act – Stop Legii Pirateriei Online) din data de 18 ianuarie 2012 a fost atât de însemnată: am avut sentimentul că o parte din creierul meu se afla în grevă (cu toate că eu și alții am găsit totuși modalități de a accesa resursele online). A mai fost și o demonstrație impresionantă a forței politice a acestor site-uri: s-a renunțat imediat la respectiva lege – care părea că se află în linie dreaptă pentru ratificare. Dar, fapt mai important, evenimentul a arătat cât de amplu ne-am externalizat deja părți din propria gândire în norul de computare. Deja, el face parte din ceea ce suntem. Din momentul în care vom deține în mod curent inteligență nonbiologică în creierul nostru, această amplificare – și norul la care e conectată ea – va continua să crească exponențial din punctul de vedere al capacității.

Inteligența pe care o vom crea prin intermediul copierii creierului cu ajutorul retroingineriei va avea acces la propriul cod-sursă și va putea să se îmbunătățească pe sine rapid, într-un ciclu recurent de proiectare în continuă accelerare. Deși creierul biologic uman beneficiază de

o plasticitate considerabilă, după cum am văzut, el are şi o arhitectură relativ fixă, care nu poate fi modificată în mod semnificativ, plus o capacitate limitată. Nu suntem capabili să creştem numărul de 300 de milioane de recognizatori de modele la 400 de milioane, să spunem, dacă nu o facem nonbiologic. Din momentul în care vom putea realiza aşa ceva, nu va mai exista niciun motiv pentru care să ne oprim la un anumit nivel de capacitate. Putem continua până la a face un miliard de recognizatori de modele sau o mie de miliarde.

Din perfecționarea cantitativă rezultă progresul calitativ. Cel mai important avans al evoluției în cazul lui *Homo sapiens* a fost cantitativ: dezvoltarea unei frunți mai mari, care să adăpostească un neocortex mai mare. Capacitatea sporită a neocortexului a permis acestei noi specii să creeze şi să cultive gânduri la niveluri conceptuale superioare, fapt concretizat în consacrarea tuturor variatelor domenii ale artei şi științei. Cu cât adăugăm mai mult neocortex la o formă nonbiologică, cu atât ne putem aştepta la niveluri calitative şi mai înalte ale abstracției.

Matematicianul britanic Irvin J. Good, coleg cu Alan Turing, a scris în anul 1965 că „prima maşină ultrainteligentă este ultima invenție pe care trebuie s-o mai facă omul". El a definit o asemenea maşină ca fiind una ce poate să depăşească „activitățile intelectuale ale oricărui om, indiferent de cât de inteligent e acesta" şi a concluzionat că „de vreme ce crearea maşinilor este una dintre aceste activități intelectuale, o maşină inteligentă ar putea să proiecteze maşini şi mai inteligente; pe urmă, se va produce fără îndoială o «explozie de inteligență»".

Ultima invenție pe care a trebuit s-o facă evoluția biologică – neocortexul – conduce inevitabil la ultima invenție pe care trebuie s-o facă omenirea – maşinile cu adevărat inteligente –, iar proiectul uneia inspiră proiectul celeilalte. Evoluția biologică mai continuă, însă evoluția tehnologică se produce de un milion de ori mai rapid decât aceasta. Potrivit Legii accelerării veniturilor, până la sfârşitul acestui secol, vom putea să creăm computație la limitele posibilului, pe baza legilor fizicii, aşa cum sunt ele aplicate computației.[5] Materia şi energia organizate astfel sunt numite „computronium" şi formează o combinație mult mai puternică sub toate aspectele decât creierul uman. Nu va fi vorba doar de computare brută, aceasta urmând să fie infuzată cu algoritmii inteligenți care constituie întreaga cunoaştere om-maşină. În timp, vom converti mare parte din masa şi energia micului nostru colț de galaxie, care este adecvat pentru acest scop, în computronium. Pe urmă,

pentru a ține în stare de funcțiune Legea accelerării veniturilor, va trebui să ne răspândim în restul galaxiei şi al Universului.

Dacă viteza luminii rămâne într-adevăr o limită inexorabilă, colonizarea Universului va dura foarte mult timp, dat fiind faptul că sistemul solar cel mai apropiat de Pământ se află la o distanță de patru ani-lumină de acesta. Dacă există fie şi cele mai subtile mijloace de ocolire a acestei limite, inteligența şi tehnologia noastră vor fi suficient de puternice pentru a le exploata. Acesta este unul dintre motivele pentru care recenta sugestie că miuonii care au străbătut 730 de kilometri de la acceleratorul CERN aflat la frontiera franco-elvețiană până în laboratorul Gran Sasso din centrul Italiei şi au părut să se deplaseze mai rapid decât viteza luminii au constituit o ştire potențial importantă. Această observație pare să fie o falsă alarmă, dar există şi alte posibilități de ocolire a acestei limite. Nici măcar nu trebuie să depăşim viteza luminii dacă putem identifica scurtături către locuri aparent depărtate prin alte dimensiuni ale spațiului în afara celor trei cu care suntem familiarizați. Dacă ne dovedim sau nu capabili să depăşim ori să ocolim în vreun fel viteza luminii ca limită, asta va fi o chestiune strategică pentru civilizația om-maşină de la începutul secolului douăzeci şi doi.

În rândurile cosmologilor persistă o controversă pe tema sfârşitului lumii în flăcări (Big Crunch, care să corespundă începutului, Big Bang) sau în gheață (moartea stelelor, pe măsură ce se depărtează într-o permanentă expansiune), dar nu se ia în calcul forța inteligenței, ca şi cum apariția ei n-ar fi fost decât un spectacol de divertisment colateral față de marea mecanică cerească, stăpâna actuală a Universului. De cât timp vom avea nevoie pentru a ne răspândi inteligența în forma ei nonbiologică în întregul Univers? Dacă putem depăşi viteza luminii – recunosc, un uriaş „dacă" –, de exemplu, folosind găurile de vierme din spațiu (care sunt compatibile cu modul actual în care înțelegem noi fizica), asta s-ar putea realiza în decurs de câteva secole. Altminteri, e posibil să dureze cu mult mai mult. Indiferent de scenariu, destinul nostru este trezirea Universului şi apoi deciderea conştientă a sorții sale insuflându-i inteligența noastră umană în forma ei nonbiologică.

NOTE

Introducere

1. Iată o frază din *Un veac de singurătate* de Gabriel García Márquez: „Aureliano Segundo nu băgă de seamă la această poliloghie de reproşuri până a doua zi după micul dejun când se simţi sâcâit de un bâzâit care devenise deja mai fluid şi mai zgomotos decât gălăgia ploii şi era Fernanda care umbla prin casă, lamentându-se că fusese crescută să fie regină şi ajunsese slujnică într-o casă de nebuni, cu un soţ puturos, idolatru şi libertin care aştepta să-i pice pâinea în gură din cer în timp ce ea se spetea să ţină, cu chiu, cu vai pe linia de plutire o casă în care erau atât de multe de făcut, atâtea de îndurat şi atâtea de reparat de când făcea Bunul Dumnezeu lumină pe pământ în zori şi până venea vremea de culcare, că atunci când se întindea o usturau ochii de parcă ar fi avut sticlă pisată sub pleoape şi nimeni nu-i zicea şi ei Bună dimineaţa, *Fernandina, ai dormit bine?* şi nici n-o întreba careva, măcar de politeţe, de ce era atât de palidă şi de ce se scula cu cearcănele-alea violete sub ochi, deşi se aşteptase la astea, sigur că da, din partea unei familii care o considerase întotdeauna o belea, o zdreanţă, o paiaţă desenată pe perete şi care peste tot pe unde se nimerea o tot bârfea pe la spate, o făcea şoarece de biserică, farisee, vicleană, ba chiar şi Amaranta, odihnească-se în pace, zisese în gura mare că era una de-aia care-şi încurcă ţoalele cu poalele, ferească Dumnezeu, ce cuvinte!, iar ea îndurase totul resemnată mulţumită Sfântului Părinte, dar nu se mai putuse abţine când diavolul ăla de José Arcadio Segundo zisese că peste familie căzuse blestemul când îşi deschisese porţile unei muntence, auzi vorbă, unei muntence încrezute, mântuieşte-ne Doamne, odraslă de muntean al dracu' din aceeaşi stirpe cu muntenii trimişi de guvern să îi omoare pe muncitori, auzi dumneata, şi nu la alta bătea, ci chiar la ea, fina Ducelui de Alba, o doamnă de asemenea descendenţă că le dădea crize de ficat nevestelor de preşedinţi, o femeie de sânge nobil ca ea, care avea dreptul să se semneze cu unsprezece patronime peninsulare şi care era singura fiinţă muritoare din orăşelul acela plin de ticăloşi care nu era pusă deloc în încurcătură când vedea şaisprezece feluri de tacâmuri în faţa ochilor, ca să aibă apoi desfrânatul de bărbat-so de ce să se prăpădească de râs şi să spună că atâtea cuţite şi furculiţe nu erau menite unei fiinţe omeneşti, ci unui miriapod şi tot ea era şi singura care putea să spună cu ochii închişi când să se servească vinul alb, pe ce parte şi în care pahar şi când vinul roşu şi din ce parte şi în ce pahar, nu ca ţăranca aia de Amaranta, odihnească-se în pace, care credea că vinul alb se

serveşte ziua şi vinul roşu noaptea şi era singura de pe toată coasta care se putea mândri că se îngrijea de nevoile ei trupeşti numai în oale de noapte din aur, aşa că şi colonelul Aureliano Buendía, odihneaşcă-se în pace, să aibă îndrăzneala să o întrebe, cu umorul lui rău de mason, de unde îşi căpătase privilegiul şi dacă de răhăţit se răhăţea tot rahat sau nu cumva de-a dreptul busuioc, auzi vorbă, chiar în cuvintele astea, şi tocmai Renata, fiica ei, care dintr-o neatenţie o văzuse făcându-şi treaba mare în dormitor, să-i răspundă că chiar dacă oala era din aur şi cu însemne heraldice, ce era înăuntru tot rahat era, rahat de-a dreptul şi mai rău chiar decât oricare altul pentru că era rahat vârtos de munteancă, auzi vorbă de la propria ei fiică, aşa că nu-şi mai făcuse niciun fel de iluzii în privinţa restului familiei, dar, în orice caz, avea dreptul să se aştepte la ceva mai multă consideraţiune din partea soţului pentru că, de bine, de rău, era soţul ei cu cununie, consortul ei, sprijinul, defloratorul ei legal, care luase asupra lui de bună voie şi nesilit de nimeni grava răspundere să o ducă din casa părintească, unde nu-i lipsise şi nu se plânsese niciodată de nimic, unde împletea cununi funerare să-şi mai treacă timpul, pentru că naşul ei trimisese o scrisoare care îi purta semnătura şi pecetea de pe inelul lui în sigiliul din ceară ca să spună că mâinile finei lui nu erau făcute pentru corvezi pe lumea asta altele decât cântatul la clavir şi, cu toate astea, nebunul de soţ o luase din casa ei cu tot felul de dojeni şi atenţionări şi o adusese în tigaia infernală în care nu putea omu' să respire de zăduf şi până să-şi încheie ea postul de Rusalii, îşi şi luase tălpăşiţa doar în nădragi şi cu acordeonul ăla al lui de chefliu să se destrăbălează în adulter cu o mizerabilă pe care era de-ajuns să o vezi din spate, aşa se spunea, cel puţin, era de-ajuns să-i vezi fundu' de iapă ca să-ţi dai seama că era o de-aia, o de-aia... de-a dreptul invers decât ea, care era o doamnă şi la castel şi în cocină, şi la masă şi în pat, doamnă din născare, cu frica lui Dumnezeu, care Îi asculta poruncile şi se supunea dorinţelor Lui şi cu care sigur că el nu putea să-şi facă toate mendrele şi porcăriile pe care le făcea cu cealaltă care, sigur, era dispusă la orice, ca matroanele franţuzoaice şi mai rea decât ele dacă stai să te gândeşti, pentru că alea cel puţin aveau obrazul să-şi pună felinar roşu la poartă, asemenea porcării, auzi dumneata şi numai asta îi mai lipsea unicei şi iubitei fiice a Doñei Renata Argote şi a lui Don Fernando del Carpio, mai ales a acestuia din urmă, un om cinstit, bun creştin, cavaler al Ordinului Sfântului Mormânt, unul din aceia care primiseră de-a dreptul de la Dumnezeu privilegiul să nu putrezească în groapă, cu pielea curată şi strălucitoare ca obrajii miresei şi limpede ca smaraldul."

2. Vezi graficul **Creşterea volumului de date despre secvenţele ADN în Genbank** de la Capitolul 10.

3. Cheng Zhang şi Jianpeng Ma, „Enhanced Sampling and Applications in Protein Folding in Explicit Solvent", *Journal of Chemical Physics* 132, nr. 24 (2010): 244101. Vezi de asemenea http://folding.stanford.edu/English/AboutaboutttheFolding@homeproject, care a folosit un număr de peste cinci milioane de computere din întreaga lume pentru simularea conformării proteinelor.

4. Pentru o descriere mai amplă a acestei argumentații, vezi secțiunea „[The Impact...] on the Intelligent Destiny of the Cosmos: Why We Are Probably Alone in the Universe" din Capitolul 6 al cărții *The Singularity Is Near* de Ray Kurzweil (New York: Viking, 2005).

5. James D. Watson, *Discovering the Brain* (Washington, DC: National Academies Press, 1992).

6. Sebastian Seung, *Connectome: How the Brain's Wiring Makes Us Who We Are* (New York: Houghton Mifflin Harcourt, 2012).

7. „Mandelbrot Zoom", http://www.youtube.com/watch?v=gEw8xpb1aRA; „Fractal Zoom Mandelbrot Corner", http://www.youtube.com/watch?v=G_GBwuYuOOs.

Capitolul 1. Experimente mintale despre lume

1. Charles Darwin, *Originea speciilor* (P.F. Collier & Son, 1909), 185/95–96.

2. Darwin, *Originea speciilor*, 751 (206.1.1-6), ediția Peckham's Variorum, editată de Morse Peckham, *The Origin of Species by Charles Darwin: A Variorum Text* (Philadelphia: University of Pennsylvania Press, 1959).

3. R. Dahm, „Discovering DNA: Friedrich Miescher and the Early Years of Nucleic Acid Research", *Human Genetics* 122, nr. 6 (2008): pp. 565–81, doi:10.1007/s00439-007-0433-0; PMID17901982.

4. Valery N. Soyfer, „The Consequences of Political Dictatorship for Russian Science", *Nature Reviews Genetics* 2, nr. 9 (2001): pp. 723– 29, doi:10. 1038/35088598; PMID 11533721.

5. J.D. Watson și F.H.C. Crick, „A Structure for Deoxyribose Nucleic Acid", *Nature* 171 (1953): 737–38, http://www.nature.com/nature/dna50/watsoncrick.pdf și „Double Helix: 50 Years of DNA", *Nature Archive*, http://www.nature.com /nature/dna50/archive.html.

6. Franklin a murit în anul 1958, iar Premiul Nobel pentru descoperirea ADN-ului a fost conferit în anul 1962. Există o controversă pe tema împărțirii acestui premiu, dacă Franklin ar mai fi trăit în 1962.

7. Albert Einstein, „On the Electrodynamics of Moving Bodies" (1905). Această lucrare a consacrat Teoria specială a relativității. Vezi Robert Bruce Lindsay și Henry Margenau, *Foundations of Physics* (Woodbridge, CT: Ox Bow Press, 1981), pag. 330.

8. „Crookes radiometer", Wikipedia, http://en.wikipedia.org/wiki/Crookes_radiometer.

9. Observă că o parte din impulsul fotonilor este transferată moleculelor de aer din incintă (deoarece nu conține un vid perfect) și apoi e transmisă de către moleculele de aer încălzite palelor morișții.

10. Albert Einstein, „Does the Inertia of a Body Depend Upon Its Energy Content?" (1905). Acecastă lucrare a consacrat celebra formulă a lui Einstein, $E = mc^2$.

11. „Albert Einstein's Letters to President Franklin Delano Roosevelt", http://hypertextbook.com/eworld/einstein.shtml.

Capitolul 3. Un model de neocortex: Teoria minții bazate pe recunoaștere

1. Unele animale nemamifere, cum ar fi ciorile, papagalii şi caracatițele, sunt creditate cu un anumit nivel de rațiune, însă una limitată şi incapabilă să creeze unelte care să aibă propriul curs evolutiv de dezvoltare. Este posibil ca aceste animale să-şi fi adaptat alte regiuni ale creierului pentru realizarea unui număr mic de nivele de gândire ierarhică, însă, pentru gândirea relativ neîngrădită, ierarhică de care e capabil omul este necesar un neocortex.

2. V.B. Mountcastle, „An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System" (1978), în Gerald M. Edelman şi Vernon B. Mountcastle, *The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function* (Cambridge, MA: MIT Press, 1982).

3. Herbert A. Simon, „The Organization of Complex Systems", în Howard H. Pattee, ed., *Hierarchy Theory: The Challenge of Complex Systems* (New York: George Braziller, Inc., 1973), http://blog.santafe.edu/wp-content/uploads/2009/03/simon1973.pdf.

4. Marc D. Hauser, Noam Chomsky şi W. Tecumseh Fitch, „The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?", *Science* 298 (noiembrie 2002): 1569–79, http://www.sciencemag.org/content/298/5598/1569.short.

5. Următorul pasaj din cartea *Transcend: Nine Steps to Living Well Forever*, de Ray Kurzweil şi Terry Grossman (New York: Rodale, 2009), descrie mai detaliat această tehnică de vis lucid: am pus la punct o metodă de rezolvare a problemelor în timpul somnului. Am perfecționat-o pentru mine pe parcursul mai multor decenii şi am descoperit mijloacele subtile prin care este posibil să meargă mai bine. Încep dându-mi o problemă când mă aşez în pat. Poate să fie orice fel de problemă. Poate fi una de matematică, un lucru care să aibă legătură cu vreuna dintre invențiile mele, o chestiune de strategie în afaceri sau chiar o problemă interpersonală. Mă gândesc la problemă vreo câteva minute, dar nu încerc s-o rezolv. Asta nu m-ar face decât să îndepărtez soluționarea creativă a problemei, care urmează să-mi vină. Încerc să mă gândesc la ea. Ce ştiu despre problemă? Ce formă ar putea să îmbrace soluția? Şi pe urmă adorm. Procedând astfel, îmi pregătesc mintea inconştientă să lucreze la problemă.

Terry: Sigmund Freud a subliniat că, atunci când visăm, mulți dintre cenzorii din creierul nostru sunt relaxați, aşa că putem visa lucruri care sunt tabùuri

din punct de vedere social, cultural şi chiar sexual. Putem visa lucruri ciudate, la care nu ne-am permite să ne gândim în timpul zilei. Este cel puţin unul dintre motivele pentru care visele mele sunt ciudate.

Ray: Mai sunt şi jaluzele profesionale care împiedică omul să gândească creativ, multe dintre ele provenind din pregătirea noastră profesională, blocaje mentale cum ar fi „Nu poţi rezolva problema procesării semnalelor în acest fel" sau „Lingvistica nu este obligată să respecte asemenea reguli". Aceste premise mentale se relaxează şi ele în timpul somnului, aşa că o să visez noi moduri de rezolvare a problemelor fără să mă împovărez cu aceste constrângeri din starea de veghe.

Terry: Mai există o parte a creierului care nu munceşte când dormim: facultăţile raţionale care evaluează faptul că o idee este rezonabilă sau ba. Deci acesta este un alt motiv pentru care în vise se întâmplă lucruri ciudate sau fantastice. Când trece un elefant printr-un perete, nu suntem şocaţi de posibilitatea unui asemenea lucru. Pur şi simplu ne spunem nouă înşine, celor din vis, „În regulă, a trecut un elefant printr-un perete, ce mare lucru". Câteodată, dacă mă trezesc în puterea nopţii, constat că am visat într-un fel ciudat sau neobişnuit problema pe care mi-am dat-o spre rezolvare.

Ray: Pasul următor are loc dimineaţa, în starea indecisă dintre somn şi veghe, adeseori numită visare lucidă. În această stare, încă mai păstrez senzaţiile şi imaginile din vis, dar sunt în posesia facultăţilor raţionale. Îmi dau seama, de exemplu, că mă aflu într-un pat. Şi pot să-mi formulez gândul că am o groază de lucruri de făcut şi că ar fi mai bine să mă scol. Însă asta ar fi o greşeală. De câte ori pot, rămân în pat şi rămân în continuare în această stare de visare lucidă, deoarece ea este cheia acestei metode creative de rezolvare a problemelor. Apropo, metoda nu funcţionează dacă sună deşteptătorul.

Cititorul: Pare să fie ce-i mai bun în cele două lumi.

Ray: Exact. Mai am încă acces la gândurile din vis legate de problema pe care mi-am dat-o cu o seară înainte, însă acum sunt suficient de conştient şi raţional pentru a evalua noile idei creative care mi-au venit în minte peste noapte. Pot să le decelez pe cele logice. După 20 de minute, poate, petrecute în această stare, în mod invariabil am o perspectivă mult mai pătrunzătoare a problemei.

Am făcut şi invenţii în acest fel (şi mi-am petrecut restul zilei scriind cereri de patente), mi-am dat seama cum să organizez materialul pentru o carte cum este aceasta şi mi-au venit idei utile pentru un set variat de probleme. Dacă am de luat o decizie-cheie, întotdeauna aplic acest proces, după care este foarte posibil să am cu adevărat încredere în deciziile mele.

Cheia procesului este să-ţi laşi mintea liberă, fără prejudecăţi şi să nu-ţi faci griji în legătură cu felul cum funcţionează metoda. E opusul disciplinei

mentale. Gândeşte-te la problemă, dar pe urmă lasă ideile să treacă peste tine când adormi. Dimineaţa, lasă-ţi din nou mintea liberă când treci în revistă ideile ciudate pe care le-au generat visele. Am constatat că e o metodă nepreţuită de punere la muncă a creativităţii naturale a viselor.

Cititor: Păi, noi, ăştia, obsedaţii de muncă, acum putem să muncim şi în vis. Nu sunt sigur că soţia o să fie prea bucuroasă.

Ray: Practic, poţi să consideri că e ca şi cum ţi-ai pune visele să muncească în locul tău.

Capitolul 4. Neocortexul biologic

1. Steven Pinker, *How the Mind Works* (New York: Norton, 1997), pp. 152–53.

2. D.O. Hebb, *The Organization of Behavior* (New York: John Wiley & Sons, 1949).

3. Henry Markram şi Rodrigo Perrin, „Innate Neural Assemblies for Lego Memory", *Frontiers in Neural Circuits* 5, nr. 6 (2011).

4. Comunicare e-mail de la Henry Markram, 19 februarie, 2012.

5. Van J. Wedeen şi colab., „The Geometric Structure of the Brain Fiber Pathways", *Science* 335, nr. 6076 (30 martie, 2012).

6. Tai Sing Lee, „Computations in the Early Visual Cortex", *Journal of Physiology*, Paris 97 (2003): pp. 121–39.

7. O listă de lucrări poate fi găsită la adresa: http://cbcl.mit.edu/people/poggio/tpcv_short_pubs.pdf.

8. Daniel J. Felleman şi David C. Van Essen, „Distributed Hierarchical Processing in the Primate Cerebral Cortex", *Cerebral Cortex* 1, nr. 1 (ianuarie/februarie 1991): 1–47. O analiză convingătoare a matematicii bayesiene a comunicaţiilor de sus în jos şi de jos în sus la nivelul neocortexului e oferită de Tai Sing Lee în „Hierarchical Bayesian Inference in the Visual Cortex", *Journal of the Optical Society of America* 20, nr. 7 (iulie 2003): pp. 1434–48.

9. Uri Hasson şi colab., „A Hierarchy of Temporal Receptive Windows in Human Cortex", *Journal of Neuroscience* 28, nr. 10 (5 martie, 2008): pp. 2539–50.

10. Marina Bedny şi colab., „Language Processing in the Occipital Cortex of Congenitally Blind Adults", *Proceedings of the National Academy of Sciences* 108, nr. 11 (15 martie, 2011): pp. 4429–34.

11. Daniel E. Feldman, „Synaptic Mechanisms for Plasticity in Neocortex", *Annual Review of Neuroscience* 32 (2009): pp. 33–55.

12. Aaron C. Koralek şi colab., „Corticostriatal Plasticity Is Necessary for Learning Intentional Neuroprosthetic Skills", *Nature* 483 (15 martie, 2012): pp. 331–35.

13. Comunicare prin e-mail de la Randal Koene, ianuarie 2012.

14. Min Fu, Xinzhu Yu, Ju Lu şi Yi Zuo, „Repetitive Motor Learning Induces Coordinated Formation of Clustered Dendritic Spines in Vivo", *Nature* 483 (1 martie, 2012): pp. 92–95.

15. Dario Bonanomi şi colab., „Ret Is a Multifunctional Coreceptor That Integrates Diffusible – and Contact – Axon Guidance Signals", *Cell* 148, nr. 3 (februarie 2012): pp. 568– 82.

16. Vezi nota de final 7 de la Capitolul 11.

Capitolul 5. Creierul vechi

1. Vernon B. Mountcastle, „The View from Within: Pathways to the Study of Perception", *Johns Hopkins Medical Journal* 136 (1975): pp. 109–31.

2. B. Roska şi F. Werblin, „Vertical Interactions Across Ten Parallel, Stacked Representations in the Mammalian Retina", *Nature* 410, nr. 6828 (29 martie, 2001): pp. 583–87; „Eye Strips Images of All but Bare Essentials Before Sending Visual Information to Brain, UC Berkeley Research Shows", comunicat de presă al Universității California, Berkeley, 28 martie 2001, www.berkeley.edu/news/media/releases/2001/03/28_wers1.html.

3. Lloyd Watts, „Reverse-Engineering the Human Auditory Pathway", în J. Liu şi colab., eds., *WCCI* 2012 (Berlin: Springer-Verlag, 2012), pp. 47–59. Lloyd Watts, „Real-Time, High-Resolution Simulation of the Auditory Pathway, with Application to Cell-Phone Noise Reduction", *ISCAS* (2 iunie, 2010): pp. 3821–24. Pentru alte lucrări, vezi http://www.lloydwatts.com/publications.html.

4. Vezi Sandra Blakeslee, „Humanity? Maybe It's All in the Wiring", *New York Times*, 11 decembrie 2003, http://www.nytimes.com/2003/12/09/science/09BRAI.html.

5. T.E.J. Behrens şi colab., „Non-Invasive Mapping of Connections between Human Thalamus and Cortex Using Diffusion Imaging", *Nature Neuroscience* 6, nr. 7 (iulie 2003): pp. 750–57.

6. Timothy J. Buschman şi colab., „Neural Substrates of Cognitive Capacity Limitations", *Proceedings of the National Academy of Sciences* 108, nr. 27 (5 iulie 2011): pp. 11252– 55, http://www.pnas.org/content/108/27/11252.long.

7. Theodore W. Berger şi colab., „A Cortical Neural Prosthesis for Restoring and Enhancing Memory", *Journal of Neural Engineering* 8, nr. 4 (august 2011).

8. Funcțiile de bază sunt funcții neliniare care pot fi combinate liniar (adunându-se funcții de bază cu greutăți diferite) pentru aproximarea oricărei funcții neliniare. A. Pouget şi L.H. Snyder, „Computational Approaches to Sensorimotor Transformations", *Nature Neuroscience* 3, nr. 11 *supliment* (noiembrie 2000): pp. 1192–98.

9. J.R. Bloedel, „Functional Heterogeneity with Structural Homogeneity: How Does the Cerebellum Operate?", *Behavioral and Brain Sciences* 15, nr. 4 (1992): pp. 666–78.

10. S. Grossberg şi R.W. Paine, „A Neural Model of Cortico-Cerebellar Interactions during Attentive Imitation and Predictive Learning of Sequential Handwriting Movements", *Neural Networks* 13, nr. 8–9 (octombrie–noiembrie 2000): pp. 999–1046.

11. Javier F. Medina şi Michael D. Mauk, „Computer Simulation of Cerebellar Information Processing", *Nature Neuroscience* 3 (noiembrie 2000): pp. 1205–11.

12. James Olds, „Pleasure Centers in the Brain", *Scientific American* (octombrie 1956): pp. 105–16. Aryeh Routtenberg, „The Reward System of the Brain", *Scientific American* 239 (noiembrie 1978): pp. 154–64. K.C. Berridge şi M.L. Kringelbach, „Affective Neuroscience of Pleasure: Reward in Humans and Other Animals", *Psychopharmacology* 199 (2008): pp. 457–80. Morten L. Kringelbach, *The Pleasure Center: Trust Your Animal Instincts* (New York: Oxford University Press, 2009). Michael R. Liebowitz, *The Chemistry of Love* (Boston: Little, Brown, 1983). W.L. Witters şi P. Jones-Witters, *Human Sexuality: A Biological Perspective* (New York: Van Nostrand, 1980).

Capitolul 6. Abilităţile transcendentale

1. Michael Nielsen, *Reinventing Discovery: The New Era of Networked Science* (Princeton, NJ: Princeton University Press, 2012), 1–3. T. Gowers şi M. Nielsen, „Massively Collaborative Mathematics", *Nature* 461, nr. 7266 (2009): pp. 879–81. „A Combinatorial Approach to Density Hales-Jewett", *Gowers's Weblog*, http://gowers.wordpress.com/2009/02/01/a-combinatorial-approach-to-density-hales-jewett/. Michael Nielsen, „The Polymath Project: Scope of Participation", 20 martie 2009, http://michaelnielsen.org/blog/?p=584. Julie Rehmeyer, „SIAM: Massively Collaborative Mathematics", *Society for Industrial and Applied Mathematics*, 1 aprilie 2010, http://www.siam.org/news/news.php?id=1731.

2. P. Dayan şi Q.J.M. Huys, „Serotonin, Inhibition, and Negative Mood", *PLoS Computational Biology* 4, nr. 1 (2008), http://compbiol.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pcbi.0040004.

Capitolul 7. Neocortexul digital inspirat de biologie

1. Gary Cziko, *Without Miracles: Universal Selection Theory and the Second Darwinian Revolution* (Cambridge, MA: MIT Press, 1955).

2. David Dalrymple; am fost mentorul lui de când avea opt ani, în anul 1999. Poţi să-i citeşti biografia aici: http://esp.mit.edu/learn/teachers/davidad/bio.html şi http://www.brainsciences.org/Research-Team/mr-david-dalrymple.html.

3. Jonathan Fildes, „Artificial Brain «10 Years Away»", BBC News, 22 iulie 2009, http://news.bbc.co.uk/2/hi/8164060.stm. Vezi de asemenea materialul video

„Henry Markram on Simulating the Brain: The Next Decisive Years", http://www.kurzweilai.net/henry-markram-simulating-the-brain-next-decisive-years.

4. M. Mitchell Waldrop, „Computer Modelling: Brain in a Box", *Nature News*, 22 februarie 2012, http://www.nature.com/news computer-modelling-brainin-a-box-1.10066.

5. Jonah Lehrer, „Can a Thinking, Remembering, Decision-Making Biologically Accurate Brain Be Built from a Supercomputer?", *Seed*, http://seedmagazine.com/content/article/out_of_the_blue/.

6. Fildes, „Artificial Brain «10 Years Away»".

7. Vezi http://www.humanconnectomeproject.org/.

8. Anders Sandberg şi Nick Bostrom, *Whole Brain Emulation: A Roadmap, Technical Report # 2008–3 (2008)*, Future of Humanity Institute, Oxford University, www.fhi.ox.ac.uk/reports/2008-3.pdf.

9. Iată schema de bază pentru un algoritm de rețea neuronală. Sunt posibile numeroase variațiuni, iar creatorul sistemului trebuie să includă anumiți parametri şi metode critice detaliate în paginile care urmează.

Crearea unei soluții de rețea neuronală pentru o problemă presupune următorii paşi:

Definirea input-ului.

Definirea topologiei rețelei neuronale (respectiv, straturile de neuroni şi conexiunile dintre neuroni).

Pregătirea rețelei neuronale cu exemple ale problemei.

Rularea rețelei neuronale pregătite pentru rezolvarea unor noi exemple ale problemei.

Lansarea pentru public a rețelei neuronale.

Aceşti paşi (cu excepția ultimului) sunt detaliați mai jos:

Input-ul problemei

Input-ul problemei în rețeaua neuronală rezidă într-o serie de numere. Acest input poate fi:

Într-un sistem de recunoaştere vizuală a modelelor, un şir bidimensional de numere reprezentând pixelii dintr-o imagine; sau

Într-un sistem de recunoaştere auditiv (de exemplu, a vorbirii), un şir bidimensional de numere reprezentând un sunet, prima dimensiune reprezentând parametrii sunetului (de exemplu componentele frecvenței), iar a doua dimensiune – diverse momente în timp; sau

Un sistem arbitrar de recunoaştere a modelelor, un şir de numere reprezentând modelul de input.

Definirea topologiei

Pentru crearea unei rețele neuronale, arhitectura fiecărui neuron constă în:

Input-uri multiple, cu fiecare input „conectat" fie la output-ul altui neuron, fie la unul dintre numerele din input.

În general, un singur output, conectat fie la input-ul altui neuron (care se află de regulă pe un strat superior), fie la output-ul final.

Setarea primului strat de neuroni

Se creează N_0 neuroni pe primul strat. Toți aceşti neuroni sunt „conectați" cu toate căile lor de input la „puncte" (respectiv, numere) din input-ul problemei. Aceste conexiuni pot fi determinate aleatoriu sau folosindu-se un algoritm evolutiv (vezi în continuare).

Se atribuie o „tărie sinaptică" fiecărei conexiuni create. Aceste tării pot să pornească de la acelaşi nivel, pot fi atribuite aleatoriu sau pot fi determinate într-un alt mod (vezi în continuare).

Setarea straturilor suplimentare de neuroni

Se setează un total de M straturi de neuroni. Pentru fiecare strat, se setează neuronii de pe strat.

Pentru stratul *i*:

Se creează N_i neuroni pe stratul *i*. Se „conectează" fiecare dintre input-urile multiple ale fiecărui neuron la output-urile neuronilor de pe stratul 1 (vezi variațiunile care urmează).

Se atribuie o „tărie sinaptică" fiecărei conexiuni create. Aceste tării pot să pornească de la acelaşi nivel, pot fi atribuite aleatoriu sau pot fi determinate într-un alt mod (vezi în continuare).

Output-urile neuronilor de pe stratul M sunt output-urile rețelei neuronale (vezi variațiunile în continuare).

Încercările de recunoaştere

Cum funcționează fiecare neuron

Odată rețeaua setată, ea face următoarele pentru fiecare încercare de recunoaştere:

Fiecare input al neuronului care corespunde tăriei este computat prin multiplicarea output-ului altui neuron (sau a input-ului inițial) la care este conectat input-ul acestui neuron, după tăria sinaptică a acelei conexiuni.

Toate input-urile neuronului care corespund ca tărie sunt însumate.

Dacă această sumă depăşeşte pragul de excitare al acestui neuron, se consideră că neuronul emite, iar output-ul lui este 1. Altminteri, output-ul lui este 0 (vezi variațiunile în continuare).

Se fac următoarele pentru fiecare încercare de recunoaştere:

Pentru fiecare strat, de la stratul *0* la stratul *M*:

Pentru fiecare neuron de pe strat:

Se însumează input-urile care corespund ca tărie (fiecare input ce corespunde ca tărie = output-ul altui neuron [sau input-ul original] la care este conectat input-ul acestui neuron, înmulțit cu tăria sinaptică a conexiunii respective).

Dacă suma de input-uri corespunzătoare tăriei este mai mare decât pragul de excitație al acestui neuron, se setează output-ul neuronului la 1; altminteri, se setează la 0.

Pentru pregătirea rețelei neuronale:

Se rulează încercări repetate de recunoaștere pe probleme-mostră.

După fiecare încercare, se ajustează tăria sinaptică a tuturor conexiunilor interneuronale pentru îmbunătățirea performanțelor rețelei neuronale la această încercare (vezi discuția despre cum se procedează în continuare).

Se continuă această pregătire până când acuratețea rețelei neuronale nu se mai îmbunătățește (respectiv, devine asimptotică).

Decizii-cheie de concepție:

În schema simplă de mai înainte, creatorul algoritmului rețelei neuronale trebuie să determine de la bun început:

Ce reprezintă numerele de input.

Numărul de straturi de neuroni.

Numărul de neuroni de pe fiecare strat. (Nu este neapărat necesar ca toate straturile să aibă același număr de neuroni.)

Numărul de input-uri pentru fiecare neuron de pe fiecare strat. Numărul de input-uri (respectiv, de conexiuni interneuronale) poate şi el să varieze de la un neuron la altul şi de la un strat la altul.

„Cablarea" propriu-zisă (respectiv, realizarea conexiunilor). Pentru fiecare neuron de pe fiecare strat, aceasta constă într-o listă de alți neuroni, ale căror output-uri constituie input-uri pentru neuronul respectiv. Este vorba despre o parte-cheie a concepției. Există mai multe căi posibile prin care se poate face acest lucru:

(1) Rețeaua neuronală este cablată aleatoriu; ori

(2) Se folosește un algoritm evolutiv (vezi în continuare) pentru a se determina cablarea optimă; ori

(3) Se folosește judecata creatorului pentru determinarea cablării.

Tăriile sinaptice inițiale (respectiv, greutatea) fiecărei conexiuni. Există mai multe moduri posibile de realizare:

(1) Se setează tăriile sinaptice la aceeaşi valoare; ori

(2) Se setează tăriile sinaptice la valori aleatorii diferite; ori

(3) Se foloseşte un algoritm evolutiv pentru determinarea unui set optim de valori iniţiale; ori
(4) Se foloseşte judecata creatorului pentru determinarea valorilor iniţiale.
Pragul de excitaţie al fiecărui neuron.
Se determină output-ul. Output-ul poate să fie:
(1) Output-urile stratului *M* de neuroni; ori
(2) Output-ul unui singur neuron, ale cărui input-uri sunt output-urile neuronilor de pe stratul *M*; ori
(3) O funcţie (de exemplu, suma) output-urilor neuronilor de pe stratul *M*; ori
(4) O altă funcţie a output-urilor neuronilor de pe mai multe straturi.
Se determină felul cum sunt ajustate tăriile sinaptice ale tuturor conexiunilor în cursul pregătirii acestei reţele neuronale. Este o decizie-cheie de concepţie şi e subiectul unui volum impresionant de cercetări şi discuţii. Există mai multe căi posibile prin care se poate face acest lucru:
(1) La fiecare încercare de recunoaştere, se măreşte sau se reduce fiecare tărie sinaptică cu o valoare fixă (în general mică), astfel încât output-ul reţelei neuronale să corespundă mai bine răspunsului corect. Un mijloc este să se încerce şi creşterea, şi descreşterea şi să se vadă care are efectul cel mai dezirabil. Este o metodă cronofagă, astfel încât există şi alte metode pentru luarea deciziilor în sensul creşterii sau descreşterii fiecărei tării sinaptice.
(2) Există şi alte metode statistice pentru modificarea tăriilor sinaptice după fiecare încercare, astfel încât performanţa reţelei neuronale la încercarea respectivă să se apropie mai mult de răspunsul corect.
A se reţine că pregătirea reţelei neuronale dă rezultate chiar şi dacă încercările de pregătire nu sunt toate corecte. Acest fapt permite folosirea de date de pregătire din lumea reală, care pot avea o rată de erori inerentă. Una dintre cheile succesului unui sistem de recunoaştere bazat pe reţele neuronale este volumul de date folosit la pregătire. De obicei, e necesar un volum foarte substanţial pentru obţinerea unor rezultate satisfăcătoare. La fel ca în cazul studenţilor umani, volumul de timp pe care îl petrece o reţea neuronală în învăţarea lecţiilor este un factor-cheie pentru performanţa lui.

Variaţiuni
Sunt fezabile mai multe variaţiuni ale celor de mai înainte. De exemplu:
Există mai multe moduri de determinare a topologiei. În particular, cablarea interneuronală poate fi setată fie aleatoriu, fie folosindu-se un algoritm evolutiv.
Există mai multe mijloace de setare a tăriilor sinaptice iniţiale.

> Input-urile pentru neuronii de pe stratul *i* nu trebuie să vină neapărat din output-urile neuronilor de pe stratul *i – 1*. Invers, input-urile pentru

neuronii de pe fiecare strat pot să vină de la orice strat inferior sau de la orice strat în mod general.

Există mai multe moduri de determinare a output-ului final.

Metoda descrisă mai înainte se concretizează prin excitații „Totul sau nimic" (1 ori 0), situație denumită nonliniaritate. Mai există și alte funcții nonliniare care pot fi folosite. În mod curent, se folosește o funcție care ajunge de la 0 la 1 de o manieră mai rapidă dar mai treptată. De asemenea, output-urile pot fi și alte numere, nu doar 0 și 1.

Diversele metode de ajustare a tăriilor sinaptice în cursul pregătirii constituie decizii-cheie de concepție.

Schema de mai înainte reprezintă o rețea neuronală „sincronă" în care fiece recunoaștere de model se realizează prin computarea output-urilor de la fiecare nivel, începând cu stratul *O* și mergând până la stratul *M*. În cazul unui sistem paralel real, în care fiece neuron operează independent de ceilalți, neuronii pot opera „asincron" (respectiv, independent). În cazul unei abordări asincrone, fiecare neuron își scanează permanent input-urile și emite de câte ori suma input-urilor sale ce corespund tăriei îi depășește pragul (sau orice altceva ar specifica funcția sa de output).

10. Robert Mannell, „Acoustic Representations of Speech", 2008, http://clas.mq.edu.au/acoustics/frequency/acoustic_speech.html.

11. Iată schema de bază pentru un algoritm genetic (evolutiv). Sunt posibile multe variațiuni, iar creatorul sistemului trebuie să asigure anumiți parametri și metode critice, detaliate în continuare.

Algoritmul evolutiv

Se creează *N* „ființe"-soluții. Fiecare are:

Un cod genetic: o succesiune de numere ce caracterizează o anumită soluție a problemei. Numerele pot constitui parametri critici, pași către o soluție, reguli etc.

Pentru fiecare generație de evoluție se procedează după cum urmează:

Se procedează după cum urmează pentru fiecare dintre cele N ființe soluții:

Se aplică soluția ființei-soluție (așa cum este reprezentată de codul său genetic) la problemă ori la mediul simulat. Se evaluează soluția.

Se aleg *L* ființe-soluție cu evaluările cele mai bune care supraviețuiesc în generația următoare.

Se elimină *(N – L)* ființe-soluție care nu supraviețuiesc.

Se creează *(N – L)* noi ființe-soluție din cele *L* ființe-soluție care au supraviețuit, prin:

(1) Realizarea de copii ale celor *L* ființe supraviețuitoare. Se introduc mici variații aleatorii pentru fiecare copie; ori

(2) Se creează alte ființe-soluție suplimentare, prin combinarea unor părți din codul genetic (folosindu-se reproducerea „sexuală" ori combinarea unor porțiuni de cromozomi) din ființele supraviețuitoare *L*; ori

(3) Se face o combinație din (1) și (2).

Se decide dacă se continuă sau nu evoluția:

Îmbunătățire = (evaluările cele mai înalte ale generației) – (evaluările cele mai înalte ale generației anterioare).

Dacă Îmbunătățirea < Pragul de Îmbunătățire, atunci am terminat.

Ființa-soluție cu evaluarea cea mai înaltă din ultima generație de evoluție are soluția cea mai bună. Se aplică soluția definită de codul ei genetic la problemă.

Decizii-cheie de concepție

În schema simplă de mai înainte, creatorul trebuie să determine de la bun început:

Parametrii-cheie:

N

L

Pragul de îmbunătățire.

Ce reprezintă numerele din codul genetic și cum este computată soluția din acest cod.

O metodă de determinare a celor *N* ființe-soluție din prima generație. În general, acestea trebuie să fie doar tentative „rezonabile" de soluție. Dacă soluțiile din prima generație sunt prea îndepărtate, este posibil ca algoritmul evolutiv să aibă dificultăți în privința ajungerii la o soluție bună. Adeseori, merită să se creeze de la bun început ființe-soluție inițiale astfel încât să fie rezonabil de diferite. Asta ajută procesul evolutiv să nu se blocheze la o soluție optimă „locală".

Cum se evaluează soluțiile.

Cum se reproduc ființele-soluție supraviețuitoare.

Variațiuni

Sunt fezabile mai multe dintre variațiunile de mai înainte. De exemplu:

Nu trebuie să existe un număr fix de ființe-soluție supraviețuitoare (*L*) pentru fiecare generație. Regula (regulile) supraviețuirii poate (pot) să permită un număr variabil de supraviețuitori.

Nu trebuie să existe un număr fix de noi ființe-soluție în fiecare generație (*N* – *L*). Regulile de procreare pot fi independente de mărimea populației.

Procrearea poate fi asociată cu supraviețuirea, prin aceasta permițându-li-se celor mai adecvate ființe-soluție să se reproducă cel mai des.
Decizia continuării sau stopării evoluției poate fi diversificată. Ea poate avea în vedere nu numai ființele-soluție cu cele mai bune evaluări din cea (cele) mai recentă (recente) generație (generații). Poate să ia în considerare și o tendință care să meargă mai în urmă de ultimele două generații.

12. Dileep George, „How the Brain Might Work: A Hierarchical and Temporal Model for Leaming and Recognition" (dizertație de doctorat, Universitatea Stanford, iunie 2008).

13. A.M. Turing, „Computing Machinery and Intelligence", *Mind*, octombrie 1950.

14. Hugh Loebner are o competiție dotată cu „Premiul Loebner" care se desfășoară în fiecare an. Medalia de argint Loebner este acordată unui computer care trece Testul Turing original exclusiv text. Medalia de aur va fi acordată unui computer care poate să treacă o versiune a testului ce include input și output audio și video. Din punctul meu de vedere, de fapt, includerea testului audio și video nu crește dificultatea probei.

15. „Cognitive Assistant That Learns and Organizes", *Artificial Intelligence Center, SRI International*, http://www.ai.sri.com/project/CALO.

16. Dragon Go! *Nuance Communications, Inc.*, http://www.nuance.com/products /dragon-go-in-action/index.htm.

17. „Overcoming Artificial Stupidity", *WolframAlpha Blog*, 17 aprilie 2012, http://blog.wolframalpha.com/author/stephenwolfram/.

Capitolul 8. Mintea ca un computer

1. Salomon Bochner, *A Biographical Memoir of John von Neumann* (Washington, DC: National Academy of Sciences, 1958).

2. A.M. Turing, „On Computable Numbers, with an Application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society* Series 2, vol. 42 (1936–37): pp. 230–65, http://www.comlab.ox.ac.uk/activities/ieg/e-library/sources/tp2-ie.pdf. A.M. Turing, „On Computable Numbers, with an Application to the Entscheidungsproblem: A Correction", *Proceedings of the London Mathematical Society* 43 (1938): pp. 544–46.

3. John von Neumann, „First Draft of a Report on the EDVAC", Moore School of Electrical Engineering, University of Pennsylvania, 30 iunie 1945. John von Neumann, „A Mathematical Theory of Communication", *Bell System Technical Journal*, iulie și octombrie 1948.

4. Jeremy Bernstein, *The Analytical Engine: Computers-Past, Present, and Future*, ediție revăzută (New York: William Morrow & Co., 1981).

5. „Japan's K Computer Tops 10 Petaflop/s to Stay Atop TOP500 List", *Top 500*, 11 noiembrie 2011, http://top500.org/lists/2011/11/press-release.

6. Carver Mead, *Analog VLSI and Neural Systems* (Reading, MA: Addison-Wesley, 1986).

7. „IBM Unveils Cognitive Computing Chips", comunicat de presă IBM, 18 august 2011, http://www-03.ibm.com/press/us/en/pressrelease/35251.wss.

8. „Japan's K Computer Tops 10 Petaflop/s to Stay Atop TOP500 List".

Capitolul 9. Experimente mintale despre minte

1. John R. Searle, „I Married a Computer", în Jay W. Richards, ed., *Are We Spiritual Machines? Ray Kurzweil vs. the Critics of Strong AI* (Seattle: Discovery Institute, 2002).

2. Stuart Hameroff, *Ultimate Computing: Biomolecular Consciousness and Nanotechnology* (Amsterdam: Elsevier Science, 1987).

3. P.S. Sebel şi colab., „The Incidence of Awareness during Anesthesia: A Multicenter United States Study", *Anesthesia and Analgesia* 99 (2004): pp. 833–39.

4. Stuart Sutherland, *The International Dictionary of Psychology* (New York: Macmillan, 1990).

5. David Cockburn, „Human Beings and Giant Squids", *Philosophy* 69, nr. 268 (aprilie 1994): pp. 135–50.

6. Ivan Petrovici Pavlov, dintr-o prelegere ţinută în anul 1913, publicată în *Lectures on Conditioned Reflexes: Twenty-Five Years of Objective Study of the Higher Nervous Activity [Behavior] of Animals* (London: Martin Lawrence, 1928), pag. 222.

7. Roger W. Sperry, din James Arthur, *Lecture on the Evolution of the Human Brain*, 1964, pag. 2.

8. Henry Maudsley, „The Double Brain", *Mind* 14, nr. 54 (1889): pp. 161–87.

9. Susan Curtiss şi Stella de Bode, „Language after Hemispherectomy", *Brain and Cognition* 43, nr. 1–3 (iunie-august 2000): pp. 135–38.

10. E.P. Vining şi colab., „Why Would You Remove Half a Brain? The Outcome of 58 Children after Hemispherectomy – the Johns Hopkins Experience: 1968 to 1996", *Pediatrics* 100 (august 1997): pp. 163–71. M.B. Pulsifer şi colab., „The Cognitive Outcome of Hemispherectomy in 71 Children", *Epilepsia* 45, nr. 3 (martie 2004): pp. 243–54.

11. S. McClelland III şi R.E. Maxwell, „Hemispherectomy for Intractable Epilepsy in Adults: The First Reported Series", *Annals of Neurology* 61, nr. 4 (aprilie 2007): pp. 372–76.

12. Lars Muckli, Marcus J. Naumerd şi Wolf Singer, „Bilateral Visual Field Maps in a Patient with Only One Hemisphere", *Proceedings of the National Academy of Sciences* 106, nr. 31 (4 august 2009), http://dx.doi.org/10.1073/pnas.0809688106.

13. Marvin Minsky, *The Society of Mind* (New York: Simon and Schuster, 1988).

14. F. Fay Evans-Martin, *The Nervous System* (New York: Chelsea House, 2005), http://www.scribd.com/doc/5012597/The-Nervous-System.

15. Benjamin Libet, *Mind Time: The Temporal Factor in Consciousness* (Cambridge, MA: Harvard University Press, 2005).

16. Daniel C. Dennett, *Freedom Evolves* (New York: Viking, 2003).

17. Michael S. Gazzaniga, *Who's in Charge? Free Will and the Science of the Brain* (New York: Ecco/HarperCollins, 2011).

18. David Hume, *An Enquiry Concerning Human Understanding* (1765), ediția a doua, sub redacția Eric Steinberg (Indianapolis: Hackett, 1993).

19. Arthur Schopenhauer, *The Wisdom of Life.*

20. Arthur Schopenhauer, *On the Freedom of the Will* (1839).

21. Din Raymond Smullyan, *5000 B.C. and Other Philosophical Fantasies* (New York: St. Martin's Press, 1983).

22. Pentru o examinare lămuritoare şi incitantă pe teme similare legate de identitate şi conştiință, vezi Martine Rothblatt, „The Terasem Mind Uploading Experiment", *International Journal of Machine Consciousness* 4, nr. 1 (2012): pp. 141–58. În această lucrare, Rothblatt examinează tema identității sub aspectul unui software care imită o persoană în temeiul unei „baze de date de interviuri video şi informații asociate referitoare la persoana predecesoare". În cadrul acestui viitor experiment propus, software-ul imită cu succes persoana pe care o copiază.

23."How Do You Persist When Your Molecules Don't?", *Science and Consciousness Review* 1, nr. 1 (iunie 2004), http://www. sci-con.org/articles/20040601. html.

Capitolul 10. Legea accelerării veniturilor aplicată creierului

1. „DNA Sequencing Costs", *National Human Genome Research Institute*, NIH, http://www.genome.gov/sequencingcosts/.

2. „Genetic Sequence Data Bank, Distribution Release Notes", 15 decembrie 2009, *National Center for Biotechnology Information, National Library of Medicine*, ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt.

3. „DNA Sequencing – The History of DNA Sequencing", 2 ianuarie 2012, http://www.dnasequencing.org/history-of-dna.

4. „Cooper's Law", *ArrayComm,* http://www.arraycomm.com/technology/coopers-law.

5. „The Zettabyte Era", *Cisco*, http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.html şi „Number of Internet Hosts", *Internet Systems Consortium*, http://www.isc.org/solutions /survey/history.

6. TeleGeography © PriMetrica, Inc., 2012.

7. Dave Kristula, „The History of the Internet" (martie 1997, actualizat august 2001), „Internet Timeline v8. 0" http://www.zakon.org/robert/internet/timeline; Quest Communications, 8-K for 9/13/1998 EX-99.1; Converge! *Network Digest*, 5 decembrie 2002, http://www.convergedigest.com/Daily/daily.asp?vn=v9n229&fecha=December%2005,%202002; Jim Duffy, „AT&T Plans Backbone Upgrade to 40G", *Computerworld*, 7 iunie 2006, http://www.computerworld.com/action/article 'history.shtml;RobertZakon,"Hobbes net- www.davesite.com/webstation/http://.do?command=viewArticleBasic&articleId=9001032; „40G: The Fastest Connection You Can Get?" *InternetNews. com*, 2 noiembrie 2007, http://www.internetnews.com/infra/article.php/3708936; „Verizon First Global Service Provider to Deploy 100G on U.S. Long-Haul Network", comunicat de presă, *Verizon*, http://newscenter.verizon.com/press-releases/verizon/2011/verizon-first-global-service.html.

8. Facebook, „Key Facts", http://newsroom.fb.com/content/default.aspx?NewsAreaId=22.

9. http://www.kurzweilai.net/how-my-predictions-are-faring.

10. Calcule pe secundă la 1.000 de dolari.

Anul	*Calcule pe sec. la 1.000 de dolari*	*Maşină*	*Logaritm natural*
1900	5.82E-06	Analytical Engine	-12.05404
1908	1.30E-04	Hollerith Tabulator	-8.948746
1911	5.79E-05	Monroe Calculator	-9.757311
1919	1.06E-03	IBM Tabulator	-6.84572
1928	6.99E-04	National Ellis 3000	-7.265431
1939	8.55E-03	Zuse 2	-4.762175
1940	1.43E-02	Bell Calculator Model	-4.246797
1941	4.63E-02	Zuse 3	-3.072613
1943	5.31E+00	Colossus	1.6692151
1946	7.98E-01	ENIAC	-0.225521
1948	3.70E-01	IBM SSEC	-0.994793
1949	1. 84E+00	BINAC	0.6081338
1949	1. 04E+00	EDSAC	0.0430595
1951	1. 43E+00	Univac I	0.3576744
1953	6.10E+ 00	Univac 1103	1.8089443
1953	1.19E+ 01	IBM 701	2.4748563
1954	3.67E-01	EDVAC	-1.002666
1955	1.65E+01	Whirlwind	2.8003255
1955	3.44E+00	IBM 704	1.2348899

1958	3.26E-01	Datamatic 1000	-1.121779
1958	9.14E-01	Univac II	0.089487
1960	1.51E+00	IBM 1620	0.4147552
1960	1.52E+02	DEC PDP	-15.0205856
1961	2.83E+02	DEC PDP	-45.6436786
1962	2.94E+01	Univac III	3.3820146
1964	1.59E+02	CDC 6600	5.0663853
1965	4.83E+02	IBM 1130	6.1791882
1965	1.79E+03	DEC PDP	-87.4910876
1966	4.97E+01	IBM 360 Model 75	3.9064073
1968	2.14E+02	DEC PDP-10	5.3641051
1973	7.29E+02	Intellec8	6.5911249
1973	3.40E+03	Data General Nova	8.1318248
1975	1.06E+04	Altair 8800	9.2667207
1976	7.77E+02	DEC PDP-11 Model 70	6.6554404
1977	3.72E+ 03	Cray 1	8.2214789
1977	2.69E+04	Apple II	10.198766
1979	1.11E+03	DEC VAX 11 Model 780	7.0157124
1980	5.62E+03	Sun-1	8.6342649
1982	1.27E+05	IBM PC	11.748788
1982	1.27E+05	Compaq Portable	11.748788
1983	8.63E+04	IBM AT-80286	11.365353
1984	8.50E+04	Apple Macintosh	11.350759
1986	5.38E+05	Compaq Deskpro 386	13.195986
1987	2.33E+05	Apple Mac II	12.357076
1993	3.55E+06	Pentium PC	15.082176
1996	4.81E+07	Pentium PC	17.688377
1998	1.33E+08	Pentium II PC	18.708113
1999	7.03E+08	Pentium III PC	20.370867
2000	1.09E+08	IBM ASCI White	18.506858
2000	3.40E+08	Power Macintosh G4/500	19.644456
2003	2.07E+09	Power Macintosh G5 2.0	21.450814
2004	3.49E+09	Dell Dimension 8400	21.973168
2005	6.36E+09	Power Mac G5 Quad	22.573294
2008	3.50E+10	Dell XPS 630	24.278614
2008	2.07E+10 Mac Pro		23.7534
2009	1.63E+10 Intel Core i7 Desktop		23.514431
2010	5.32E+10 Intel Core i7 Desktop		24.697324

11. Top 500 Supercomputer Sites, http://top500.org/.

12. „Microprocessor Quick Reference Guide", Intel Research, http://www.intel.com/pressroom/kits/quickreffam.htm.

13. 1971–2000: VLSI Research Inc.

2001–2006: *The International Technology Roadmap for Semiconductors*, 2002 Update and 2004 Update, Table 7a, „Cost—Near-term Years", „DRAM cost/bit at (packaged microcents) at production".

2007–2008: *The International Technology Roadmap for Semiconductors*, 2007, Tables 7a and 7b, " Cost—Near-term Years", „Cost—Long-term Years", http://www.itrs.net/Links/2007ITRS/ExecSum2007.pdf.

2009–2022: *The International Technology Roadmap for Semiconductors*, 2009, Tables 7a and 7b, „Cost—Near-term Years", „Cost—Long-term Years", http://www.itrs.net/Links/2009ITRS/Home2009.htm.

14. Pentru a face comparabilă valoarea dată în dolari, prețurile computerelor pentru toți anii au fost convertite în echivalentul lor din anul 2000, folosindu-se datele Federal Reserve Board's CPI de la adresa http://minneapolisfed.org/research/data/us/calc/. De exemplu, un milion de dolari din anul 1960 este echivalent cu 5,8 milioane de dolari în anul 2000, iar un milion de dolari din anul 2004 este echivalent cu 0,91 milioane de dolari din anul 2000.

1949: http://www.cl.cam.ac.uk/UoCCL/misc/EDSAC99/statistics.html, http://www.davros.org/misc/chronology.html.

1951: Richard E. Matick, *Computer Storage Systems and Technology* (New York: John Wiley & Sons, 1977); http://inventors.about.com/library/weekly/aa062398.htm.

1955: Matick, *Computer Storage Systems and Technology*; OECD, 1968, http://members.iinet.net.au/~dgreen/timeline.html.

1960: ftp://rtfm.mit.edu/pub/usenet/alt.sys.pdp8/PDP-8_Frequently_Asked_Questions_%28posted_every_other_month%29; http:// www.dbit.com/~greeng3/pdp1/pdp1.html#INTRODUCTION.

1962: ftp://rtfm.mit.edu/pub/usenet/alt.sys.pdp8/PDP-8_Frequently_Asked_Questions_%28posted_every_other_ month%29.

1964: Matick, *Computer Storage Systems and Technology*; http://www.research.microsoft.com/users/gbell/craytalk; http://www.ddj.com/documents/s=1493/ddj0005hc/.

1965: Matick, *Computer Storage Systems and Technology*; http://www.fourmilab.ch/documents/univac/config1108.html; http://www.frobenius.com/univac.htm.

1968: Data General.

1969, 1970: http://www.eetimes.com/special/special_issues/millennium/milestones/whittier.html.

1974: Scientific Electronic Biological Computer Consulting (SCELBI).
1975–1996: publicitate din revista *Byte*.
1997–2000: publicitate din revista *PC Computing*.
2001: www.pricewatch.com (http://www.jc-news.com/parse.cgi?news/pricewatch/raw/pw-010702).
2002: www.pricewatch.com (http://www.jc-news.com/parse.cgi?news/pricewatch/raw/pw-020624).
2003: http://sharkyextreme.com/guides/WMPG/article php/10706_2227191_2.
2004: http://www.pricewatch.com (11/17/04).
2008: http://www.pricewatch.com (10/02/08) ($16.61).

15. Dataquest/Intel şi Pathfinder Research:

Anul	*Dolari*	*Logaritm natural*
1968	1.00000000	0
1969	0.85000000	-0.16252
1970	0.60000000	-0.51083
1971	0.30000000	-1.20397
1972	0.15000000	-1.89712
1973	0.10000000	-2.30259
1974	0.07000000	-2.65926
1975	0.02800000	-3.57555
1976	0.01500000	-4.19971
1977	0.00800000	-4.82831
1978	0.00500000	-5.29832
1979	0.00200000	-6.21461
1980	0.00130000	-6.64539
1981	0.00082000	-7.10621
1982	0.00040000	-7.82405
1983	0.00032000	-8.04719
1984	0.00032000	-8.04719
1985	0.00015000	-8.80488
1986	0.00009000	-9.31570
1987	0.00008100	-9.42106
1988	0.00006000	-9.72117
1989	0.00003500	-10.2602
1990	0.00002000	-10.8198
1991	0.00001700	-10.9823
1992	0.00001000	-11.5129

1993	0.00000900	-11.6183
1994	0.00000800	-11.7361
1995	0.00000700	-11.8696
1996	0.00000500	-12.2061
1997	0.00000300	-12.7169
1998	0.00000140	-13.4790
1999	0.00000095	-13.8668
2000	0.00000080	-14.0387
2001	0.00000035	-14.8653
2002	0.00000026	-15.1626
2003	0.00000017	-15.5875
2004	0.00000012	-15.9358
2005	0.000000081	-16.3288
2006	0.000000063	-16.5801
2007	0.000000024	-17.5452
2008	0.000000016	-17.9507

16. Steve Cullen, In-Stat, septembrie 2008, www.instat.com.

Anul	*Mbiți*	*Biți*
1971	921.6	9.216E+08
1972	3788.8	3.789E+09
1973	8294.4	8.294E+09
1974	19865.6	1.987E+10
1975	42700.8	4.270E+10
1976	130662.4	1.307E+11
1977	276070.4	2.761E+11
1978	663859.2	6.639E+11
1979	1438720.0	1.439E+12
1980	3172761.6	3.173E+12
1981	4512665.6	4.513E+12
1982	11520409.6	1.152E+13
1983	29648486.4	2.965E+13
1984	68418764.8	6.842E+13
1985	87518412.8	8.752E+13
1986	192407142.4	1.924E+14
1987	255608422.4	2.556E+14
1988	429404979.2	4.294E+14
1989	631957094.4	6.320E+14

1990	950593126.4	9.506E+14
1991	1546590618	1.547E+15
1992	2845638656	2.846E+15
1993	4177959322	4.178E+15
1994	7510805709	7.511E+15
1995	13010599936	1.301E+16
1996	23359078007	2.336E+16
1997	45653879161	4.565E+16
1998	85176878105	8.518E+16
1999	1.47327E+11	1.473E+17
2000	2.63636E+11	2.636E+17
2001	4.19672E+11	4.197E+17
2002	5.90009E+11	5.900E+17
2003	8.23015E+11	8.230E+17
2004	1.32133E+12	1.321E+18
2005	1.9946E+12	1.995E+18
2006	2.94507E+12	2.945E+18
2007	5.62814E+12	5.628E+18

17. „Historical Notes about the Cost of Hard Drive Storage Space", http://www.littletechshoppe.com/ns1625/winchest.html; publicitate din revista *Byte*, 1977–1998; publicitate din revista *PC Computing*, 3/1999; *Understanding Computers: Memory and Storage* (New York: Time Life, 1990); http://www.cedmagic.com/history/ibm-305-ramac.html; John C. McCallum, „Disk Drive Prices (1955–2012)", http://www.jcmit.com/diskprice.htm; IBM, „Frequently Asked Questions," http://www-03.ibm.com/ibm/history/documents/pdf/faq.pdf; IBM, „IBM 355 Disk Storage Unit", http://www-03.ibm.com/ibm/history/exhibits/storage/storage_355.html; IBM, „IBM 3380 Direct Access Storage Device," http://www.03-ibm. com/ibm/history/exhibits/storage/storage_3380.html.

18. „Without Driver or Map, Vans Go from Italy to China", *Sydney Morning Herald*, 29 octombrie 2010, http://www.smh.com.au/technology/technology-news/without-driver-or-map-vans-go-from-italy-to-china-20101029-176ja.html.

19. KurzweilAI.net.

20. Material adaptat cu permisiunea lui Amiram Grinvald şi Rina Hildesheim, „VSDI: A New Era in Functional Imaging of Cortical Dynamics", *Nature Reviews Neuroscience 5* (noiembrie 2004), pp. 874–85. Principalele instrumente pentru obținerea de imagini ale creierului sunt expuse în această diagramă. Capacitățile lor sunt reprezentate de dreptunghiurile haşurate. Rezoluția spațială se referă la cea mai mică dimensiune care poate fi măsurată prin intermediul unei tehnici. Rezoluția temporală reprezintă durata imaginii. În cazul fiecărei tehnici, se fac concesii. De

exemplu, EEG (electroencefalografia) care măsoară „undele cerebrale" (semnalele electrice ale neuronilor) poate măsura unde cerebrale foarte rapide (care se produc la intervale foarte apropiate), însă percepe numai semnalele din apropierea suprafeței creierului.

În schimb, rezonanța magnetică funcțională fRMN, care folosește o mașină specială pentru măsurarea fluxului de sânge la nivelul neuronilor (indicând activitatea acestora), poate să simtă mult mai profund în creier (și la nivelul măduvei spinării) și cu o rezoluție superioară, de până la câteva zeci de microni (milionimi de metru). Însă fRMN funcționează foarte lent comparativ cu EEG.

Sunt tehnici neinvazive (nu necesită intervenție chirurgicală sau administrarea de medicamente). MEG (magnetoencefalografia) este o altă tehnică neinvazivă. Detectează câmpurile magnetice generate de neuroni. MEG și EEG pot să rezolve evenimente cu rezoluție temporală de până la o milisecundă, dar mai bine decât fRMN, care poate cel mult să rezolve evenimente cu o rezoluție de câteva sute de milisecunde. De asemenea, MEG identifică precis sursele din ariile primare senzoriale somatice și motorii.

Imagistica optică acoperă aproape întreaga gamă de rezoluții spațiale și temporale, dar este invazivă. Metoda indicatorilor sensibili la tensiuni electrice este cea mai potrivită metodă de măsurare a activității cerebrale, însă, la animale, se limitează la măsurătorile realizate imediat sub suprafața cortexului.

Cortexul expus este acoperit cu o incintă transparentă sigilată; după ce e imprimat cu un indicator sensibil la tensiuni electrice, este iluminat și se imortalizează serii de fotografii cu o cameră de mare viteză. Alte tehnici optice folosite în laborator includ imagistica pe bază de ioni (de obicei de calciu sau sodiu) și sistemele de imagine cu fluorescență (imagistica multifoton și cea confocală).

Alte tehnici de laborator includ TEP (tomografia cu emisie de pozitroni, o tehnică imagistică de medicină nucleară care produce o imagine 3D), 2 DG (analiza postmortem cu 2-deoxiglucoză sau analiza tisulară), leziunile (presupune lezarea unor neuroni la animale și observarea efectelor), clamparea pe bucăți (pentru măsurarea curenților ionici pe membranele biologice) și microscopia electronică (utilizarea unui fascicul de electroni pentru examinarea țesuturilor la o scară foarte fină). Aceste tehnici pot fi de asemenea integrate în imagistica optică.

21. Rezoluția spațială RMN în microni (μm), 1980–2012:

Anul	*Rezoluția în microni*	*Citat*	*URL*
2012	125	„Characterization of Cerebral White Matter Properties Using	http://dx.doi.org/10.1089/brain.2011.0071

		Quantitative Magnetic Resonance Imaging Stains“	http://dx.doi.org/10.1089/brain.2011.0071
2010	200	„Study of Brain Anatomy with High-Field MRI: Recent Progress“	http://dx.doi.org/10.1016/j.mri.2010.02.007
2010	250	„High-Resolution Phased-Array MRI of the Human Brain at 7 Tesla: Initial Experience in Multiple Sclerosis Patients“	http://dx.doi.org/10.1111/j.1552-6569.2008.00338.x
1994	1.000	„Mapping Human Brain Activity in Vivo“	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1011409/
1989	1.700	„Neuroimaging in Patients with Seizures of Probable Frontal Lobe Origin“	http://dx.doi.org/10.1111/j.1528-157.1989.tb05470.x
1985	1.700	„A Study of the Septum Pellucidum and Corpus Callosum in Schizophrenia with MR Imaging“	http://dx.doi.org/10.1111/j.1600-0447.1985.tb02634.x
1983	1.700	„Clinical Efficiency of Nuclear Magnetic Resonance Imaging“	http://radiology.rsna.org/content/146/1/123.short
1980	5.000	„In Vivo NMR Imaging in Medicine: The Aberdeen Approach, Both Physical and Biological [and Discussion]“	http://dx.doi.org/10.1098/rstb.1980.0071

22. Rezoluția spațială în nanometri (nm) pentru tehnicile imagistice distructive, 1983– 2011:

Anul	*Rezoluția x-y (nm)*	*Citatul*	*URL*	*Tehnica*	*Note*
2011	4	„Focused Ion BeamMilling and Scanning Electron Microscopy of Brain Tissue“	http://dx.doi.org/10.3791/2588	Focused ion beam/scanning electronmicroscope (FIB/SEM)	
2011	4	„Volume ElectronMicroscopy for Neuronal Circuit Reconstruction“	http://dx.doi.org/10.1016/j.conb.2011.10.022	Scanning electronmicroscopy (SEM)	
2011	4	„Volume ElectronMicroscopy for Neuronal Circuit Reconstruction“	http://dx.doi.org/10.1016/j.conb.2011.10.022	Transmission electron microscopy (TEM)	
2004	13	„Serial Block-Face Scanning Electron Microscopy to Reconstruct Three-Dimensional Tissue Nanostructure“	http://dx.doi.org/10.1371/journal.pbio.0020329	Serial block-face scanning electron microscopy (SBF-SEM)	Rezultat citat în http://faculty.cs.tamu.edu/choe/ftp/publications/choe.hpc08-preprint.pdf, asigurat de Yoonsuck Choe.

2004	20	„Wet SEM: A Novel Method for Rapid Diagnosis of Brain Tumors“	http://dx.doi.org/10.1080/01913120490515603	„Wet“ scanning electron microscopy (wet SEM)	
1998	100	„A Depolarizing Chloride Current Contributes to Chemoelectrical Transduction in Olfactory Sensory Neurons in Situ“	http://www.jneurosci.org/content/18/17/6623.full	Scanning transmissio-nelectron microscope (STEM)	
1994	2.000	„Enhanced Optical Imaging of Rat Gliomas and Tumor Margins“	http://journals.lww.com/neurosurgery/Abstract/1994/11000/Enhanced_Optical_Imaging_of_Rat_Gliomas_and_Tumor.19.aspx	Enhanced optical imaging	With a spatial resolution of the optical images below 20 microns 2/pixel (22).
1983	3.000	„3D Imaging of X-Ray Microscopy“	http://www.scipress.org/e-library/sof2/pdf/0105.PDF	Projection microscopy	Vezi Fig. 7 din articol.

23. Rezoluția spațială în microni (μm) pentru tehnicile imagistice nedistructive la animale, 1985–2012:

Anul	*Rezoluția*	
2012	Resolution	0,07
	Citation	Sebastian Berning şi colab., „Nanoscopy in a Living Mouse Brain", *Science* 335, nr. 6068 (3 februarie 2012): 551.
	URL	http://dx.doi.org/10.1126/science.1215369
	Technique	Stimulated emission depletion (STED) fluorescence nanoscopy
	Notes	Highest resolution achieved in vivo so far
2012	Resolution	0,25
	Citation	Sebastian Berning şi colab., „Nanoscopy in a Living Mouse Brain", *Science* 335, nr. 6068 (3 februarie 2012): 551.
	URL	http://dx.doi.org/10.1126/science.1215369
	Technique	Confocal and multiphoton microscopy
2004	Resolution	50
	Citation	Amiram Grinvald şi Rina Hildesheim, „VSDI: A New Era in Functional Imaging of Cortical Dynamics", *Nature Reviews Neuroscience 5* (noiembrie 2004): pp. 874–85.
	URL	http://dx.doi.org/10.1038/nrn1536
	Technique	Imaging based on voltage-sensitive dyes (VSDI)
	Notes	„VSDI has provided high-resolution maps, which correspond to cortical columns in which spiking occurs, and offer a spatial resolution better than 50 μm."
1996	Resolution	50
	Citation	Dov Malonek şi Amiram Grinvald, „Interactions between Electrical Activity and Cortical Microcirculation Revealed by Imaging Spectroscopy: Implications for Functional Brain Mapping", *Science* 272, nr. 5261 (26 aprilie 1996): pp. 551–54.

	URL	http://dx.doi.org/10.1126/science.272.5261.551
	Technique	Imaging spectroscopy
	Notes	„The study of spatial relationships between individual cortical columns within a given brain area has become feasible with optical imaging based on intrinsic signals, at a spatial resolution of about 50 μm."
1995	Resolution	50
	Citation	D.H. Turnbull and colab., „Ultrasound Backscatter Microscope Analysis of Early Mouse Embryonic Brain Development", *Proceedings of the National Academy of Sciences* 92, nr. 6 (14 martie 1995): pp. 2239–43.
	URL	http://www.pnas.org/content/92/6/2239.short
	Technique	Ultrasound backscatter microscopy
	Notes:	„We demonstrate application of a real-time imaging method called ultrasound backscatter microscopy for visualizing mouse early embryonic neural tubes and hearts. This method was used to study live embryos in utero between 9.5 and 11.5 days of embryogenesis, with a spatial resolution close to 50 μm."
1985	Resolution	500
	Citation	H.S. Orbach, L.B. Cohen şi A. Grinvald, „Optical Mapping of Electrical Activity in Rat Somatosensory and Visual Cortex", *Journal of Neuroscience* 5, nr. 7 (01 iulie 1985): pp. 1886–95.
	URL	http://www.jneurosci.org/content/5/7/1886.short
	Technique	Optical methods

Capitolul 11. Obiecții

1. Paul G. Allen şi Mark Greaves, „Paul Allen: The Singularity Isn't Near", *Technology Review*, 12 octombrie 2011, http://www.technologyreview.com/blog/guest/27206/.

2. ITRS, „International Technology Roadmap for Semiconductors", http://www.itrs.net/Links/2011ITRS/Home2011.htm.

3. Ray Kurzweil, *The Singularity Is Near* (New York: Viking, 2005), Capitolul 2.

4. Nota de final 2 din Allen şi Greaves, „The Singularity Isn't Near", este formulată după cum urmează: „Începem să intrăm în gama de putere a computerelor de care am putea avea nevoie pentru susținerea acestui gen de simulare masivă a creierului. Computerele din clasa Petaflop (cum ar fi BlueGene/P de la IBM, care a fost folosit pentru sistemul Watson) sunt disponibile acum în comerț. Computerele din clasa Exaflop se află actualmente pe planşeta de proiectare. Aceste sisteme ar putea, probabil, că dovedească acea capacitate computațională brută necesară pentru simularea modelelor de excitație pentru toți neuronii creierului, deşi actualmente o fac de multe ori mai lent decât s-ar întâmpla într-un creier real".

5. Kurzweil, *The Singularity Is Near*, Capitolul 9, secțiunea intitulată „The Criticism from Software" (pp. 435– 42).

6. Ibid., Capitolul 9.

7. Cu toate că nu este posibil să determinăm cu precizie conținutul de informație al genomului, din cauza perechilor de baze repetate, aceasta [informația] este mult mai redusă decât datele totale necomprimate. Aici există două abordări pentru estimarea conținutului de informație comprimată din genom, ambele demonstrând că o rată de 30 la 100 de milioane de bytes este suficient de ridicată.

1. În termeni de date necomprimate, există trei miliarde de trepte de ADN în codul genetic al omului, fiecare codificând câte 2 biți (de vreme ce există patru posibilități pentru fiecare pereche de bază a ADN-ului). Prin urmare, genomul uman necomprimat este de aproximativ 800 de milioane de bytes. ADN-ul necodificat era numit odinioară „ADN junk", dar acum este clar că el joacă un rol important în expresia genelor. În orice caz, e codificat foarte ineficient. În primul rând, există redundanțe masive (de exemplu, secvența denumită „ALU" este repetată de sute de mii de ori), de care pot să profite algoritmii de compresie. Odată cu recenta explozie a băncilor de date genetice, există o doză de interes extraordinar de mare față de comprimarea datelor genetice. Cercetările recente asupra aplicării algoritmilor de date-standard pentru comprimare în cazul datelor genetice indică faptul că reducerea acestor date cu 90 la sută (la un bit de compresie perfectă) este fezabilă: Hisahiko Sato şi colab., „DNA Data Compression in the Post Genome Era", *Genome Informatics* 12 (2001): pp. 512–14, http://www.jsbi.org/journal/GIW01/GIW01P130.pdf.

Prin urmare, putem comprima genomul la aproximativ 80 de milioane de bytes fără pierdere de informație (ceea ce înseamnă că putem reconstrui perfect întregul genom necomprimat de 800 de milioane de bytes).

Acum să ne gândim că peste 98 la sută din genom nu codifică proteine. Chiar şi după o comprimare-standard a datelor (care elimină redundanțele şi foloseşte o căutare de dicționar pentru secvențele obişnuite), conținutul algoritmic al regiunilor necodificate pare să fie relativ redus, ceea ce înseamnă că e posibil să putem codifica un algoritm care ar îndeplini aceleaşi funcții cu mai puțini biți.

Pe de altă parte, de vreme ce ne aflăm abia la începutul procesului copierii genomului prin retroinginerie, nu putem face o estimare sigură a unei reduceri şi mai accentuate pe baza unui algoritm funcțional echivalent. Folosesc, prin urmare, o gamă de 30 până la 100 de milioane de bytes pentru informația comprimată din genom. Partea superioară a acestei game presupune numai comprimarea datelor, fără nicio simplificare algoritmică.

Doar o parte (deşi majoritară) a acestei informații caracterizează proiectul creierului.

2. O altă direcție de raționament este următoarea. Deşi genomul uman conține în jur de 3 miliarde de baze, doar un mic procent, cum am menționat mai înainte, codifică proteine. După estimările curente, există 26.000 de gene care codifică proteinele. Dacă presupunem că aceste gene conțin în medie 3.000 de baze de date utile, asta înseamnă doar aproximativ 78 de milioane de baze. O bază de ADN necesită numai 2 biți, fapt ce se traduce prin aproximativ 20 de milioane de bytes (78 de milioane de baze împărțit la patru). În secvența genei de codificare a proteinei, fiecare „cuvânt" (codon) de trei baze ADN se traduce printr-un aminoacid. Prin urmare, există 43 (64) de posibile coduri de codon, fiecare constând în trei baze ADN. Pe de altă parte, există numai 20 de aminoacizi folosiți, plus un codon stop (aminoacid nul) din 64. Restul de 43 de coduri sunt folosite ca sinonime ale celor 21 de coduri utile. Deci sunt necesari numai 6 biți pentru codificarea a 64 de combinații posibile şi doar aproximativ 4,4 ($\log_2 21$) biți sunt necesari pentru codificarea a 21 de posibilități, o economie de 1,6 din 6 biți (în jur de 27 la sută), ceea ce ne aduce la 15 milioane de bytes. În plus, aici este posibilă o comprimare-standard bazată pe repetarea secvențelor, deşi pe această porțiune de codificare a proteinelor e posibilă o compresie mult mai redusă decât în cazul ADN-junk, posesor al unei masive redundanțe. Aşadar, ajungem la o cifră în jur de 12 milioane de bytes. Pe de altă parte, acum trebuie să mai adăugăm informație pentru porțiunea necodificată a ADN-ului care controlează expresia genelor. Deşi această porțiune de ADN constituie masa genomului, ea pare să aibă un nivel redus de informație şi e plină de redundanțe masive. Estimând că ar corespunde unui total de

circa 12 milioane de bytes de coduri de proteine din ADN, ajungem din nou la aproximativ 24 de milioane de bytes. Din această perspectivă, o estimare de 30 până la 100 de milioane de bytes este moderat de ridicată.

8. Dharmendra S. Modha şi colab., „Cognitive Computing", *Communications of the ACM* 54, nr. 8 (2011): pp. 62–71, http://cacm.acm.org/magazines/2011/8/114944-cognitive-computing/fulltext.

9. Kurzweil, *The Singularity Is Near*, Capitolul 9, secţiunea intitulată „The Criticism from Ontology: Can a Computer Be Conscious?" (pp. 458–69).

10. Michael Denton, „Organism and Machine: The Flawed Analogy", din *Are We Spiritual Machines? Ray Kurzweil vs. the Critics of Strong AI* (Seattle: Discovery Institute, 2002).

11. Hans Moravec, *Mind Children* (Cambridge, MA: Harvard University Press, 1988).

Epilogul

1. „In U.S., Optimism about Future for Youth Reaches All-Time Low", Gallup Politics, 2 mai 2011, http://www.gallup.com/poll/147350/optimism-future-youth-reaches-time-low.aspx.

2. James C. Riley, *Rising Life Expectancy: A Global History* (Cambridge: Cambridge University Press, 2001).

3. J. Bradford DeLong, „Estimating World GDP, One Million B.C.-Present", 24 mai 1998, http://econ161.berkeley.edu/TCEH/1998_ Draft/World_GDP/Estimating_World_GDP.html şi http://futurist.typepad.com/my_weblog/2007/07/economic-growth.html. Vezi de asemenea Peter H. Diamandis şi Steven Kotler, *Abundance: The Future Is Better Than You Think* (New York: Free Press, 2012).

4. Martine Rothblatt, *Transgender to Transhuman* (publicată privat, 2011). Autoarea explică faptul că este posibil să apară o traiectorie la fel de rapidă a acceptării pentru „transumani", de exemplu, minţi nebiologice dar convingător de conştiente, după cum am discutat în Capitolul 9.

5. Următorul extras din *The Singularity Is Near*, Capitolul 3 (pp. 133–35), de Ray Kurzweil (New York: Viking, 2005), discută limitele computaţiei bazate pe legile fizicii:

> Limitele supreme ale computerelor sunt profund de înalte. Bazându-se pe lucrările profesorului Hans Bremermann de la Universitatea California din Berkeley şi ale teoreticianului nanotehnologiei Robert Freitas, profesorul Seth Lloyd de la MIT a estimat capacitatea computaţională totală, potrivit legilor cunoscute ale fizicii, a unui computer cu greutatea de un kilogram şi ocupând volumul de un litru – cam de mărimea şi greutatea unui laptop mic –, ceea ce el numeşte „laptopul suprem".

[Notă: Seth Lloyd, „Ultimate Physical Limits to Computation", *Nature* 406 (2000): pp. 1047– 54.

[Lucrările timpurii pe tema limitelor computației au fost realizate de Hans J. Bremermann în anul 1962: Hans J. Bremermann, „Optimization Through Evolution and Recombination", în M.C. Yovits, C.T. Jacobi, C.D. Goldstein, sub redacția *Self-Organizing Systems* (Washington, D.C.: Spartan Books, 1962), pp. 93–106.

[În anul 1984, Robert A. Freitas jr. a elaborat pe marginea lucrării lui Bremermann în Robert A. Freitas Jr., „Xenopsychology", *Analog* 104 (aprilie 1984): pp. 41–53, http://www.rfreitas.com/Astro/Xenopsychology.htm#SentienceQuotient.]

Volumul potențial de computație crește odată cu energia disponibilă. Putem înțelege legătura dintre energie și capacitatea computațională după cum urmează. Energia dintr-o cantitate de materie este energia asociată cu fiecare atom (și particulă subatomică). Așadar, mai mulți atomi, mai multă energie. După cum am mai spus, fiecare atom are potențialul de a fi folosit pentru computație. Mai mulți atomi, mai multă computație. Energia fiecărui atom sau particule crește odată cu frecvența mișcării: mai multă mișcare, mai multă energie. Aceeași relație există și pentru computația potențială: cu cât este mai mare frecvența mișcării, cu atât e mai mare computația pe care o poate realiza fiecare componentă (care poate fi un atom). (Vedem acest lucru în cipurile contemporane: cu cât e mai mare frecvența cipului, cu atât este mai mare viteza de computație.)

Prin urmare, există o relație direct proporțională între energia unui obiect și potențialul său de a realiza computația. Energia potențială a unui kilogram de materie este foarte mare, așa cum știm din ecuația lui Einstein $E = mc^2$. Viteza luminii reprezintă un număr foarte mare: aproximativ 10^{17} metru^2/secundă^2. Potențialul de computare al materiei este guvernat de un număr, foarte, foarte mic, constanta lui Planck: $6{,}6 \times 10^{-34}$ joule-secundă (joule-ul este o unitate de măsură pentru energie). Aceasta este cea mai mică scară la care putem aplica energia pentru computație. Obținem limita teoretică a unui obiect de a realiza computație împărțind energia totală (energia medie a fiecărui atom sau a fiecărei particule înmulțită cu numărul de asemenea particule) la constanta lui Planck.

Lloyd arată că, pentru un kilogram de materie, capacitatea potențială de computare este egală cu *pi* înmulțit cu energia și împărțit la constanta lui Planck. Deoarece energia e un număr foarte mare, iar constanta lui Planck e atât de mică, această ecuație generează un număr uriaș: în jur de 5×10^{50} operații pe secundă.

[Notă: pi × energia maximă (10^{17} kg × metru^2/secundă^2) / ($6{,}6 \times 10^{-34}$) joule-secundă) = ~ 5×10^{50} operații/secundă.]

Dacă asociem această valoare cu estimarea cea mai conservatoare a capacității creierelor umane (10^{19} cps și 10^{10} oameni), ea reprezintă echivalentul a aproximativ 5 miliarde de mii de miliarde de civilizații umane.

[Notă: 5×10^{50} cps este echivalentul a 5×10^{21} (5 miliarde de mii de miliarde) de civilizații umane (fiecare necesitând 10^{29} cps).]

Dacă folosim valoarea de 10^{16} cps, pe care o consider suficientă pentru imitarea funcțională a inteligenței umane, laptopul suprem ar funcționa cu echivalentul forței cerebrale a 5 mii de miliarde de mii de miliarde de civilizații umane.

[Notă: Zece miliarde (10^{10}) de oameni la 10^{16} cps fiecare înseamnă 10^{26} cps pentru civilizația umană. Deci 5×10^{50} cps este echivalentul a 5×10^{24} (5 mii de miliarde de mii de miliarde) de civilizații umane.]

Un asemenea laptop ar putea să realizeze echivalentul gândirii umane pe parcursul ultimelor zece mii de ani (respectiv zece miliarde de creiere omenești care operează timp de zece mii de ani) în a zece mia parte dintr-o nanosecundă.

[Notă: Această estimare emite premisa conservatoare că au existat zece miliarde de oameni în cursul ultimelor zece mii de ani, ceea ce evident că nu este cazul. Numărul real de oameni a crescut treptat în trecut pentru a ajunge la aproximativ 6,1 miliarde în anul 2000. Există 3×10^7 secunde într-un an și 3×10^{11} secunde în zece mii de ani. Așadar, folosind estimarea de 10^{26} cps pentru civilizația umană, gândirea umană pe parcursul ultimilor zece mii de ani este sigur echivalentă cu nu mai mult de 3×10^{37} calcule. Laptopul suprem realizează 5×10^{50} calcule într-o secundă. Așadar, simularea a zece mii de ani de gândire a zece miliarde de oameni ar dura în jur de 10^{-13} secunde, ceea ce înseamnă a zece mia parte dintr-o nanosecundă.]

Din nou, câteva avertismente sunt bine venite. Convertirea masei laptopului nostru de un kilogram în energie reprezintă ceea ce se întâmplă practic în cazul unei explozii termonucleare: Desigur, nu vrem ca laptopul să explodeze, ci să rămână la dimensiunile lui cu un volum în jur de un litru. Așadar, acest lucru va impune o ambalare atentă, ca să nu spunem mai mult. Analizând entropia maximă (gradele de libertate reprezentate de starea tuturor particulelor) unui asemenea dispozitiv, Lloyd arată că un laptop de felul acesta ar avea o capacitate teoretică a memoriei de 10^{31} biți. Este dificil să-ți imaginezi tehnologii care ar ajunge să atingă asemenea limite. Însă ne putem imagina foarte ușor tehnologii care se apropie rezonabil de mult. După cum arată proiectul Universității Oklahoma, am demonstrat deja capacitatea de a stoca cel puțin cincizeci de biți de informație per atom (deși numai pentru un mic număr de atomi, deocamdată). Stocarea a 10^{27} biți de memorie în cei 10^{25} atomi dintr-un kilogram de materie ar trebui, prin urmare, să poată fi realizată în cele din urmă.

Însă, deoarece multe proprietăți ale fiecărui atom pot fi exploatate pentru stocarea de informație – cum ar fi poziția precisă, spinul și starea cuantică a tuturor particulelor sale –, probabil că ne putem descurca într-un fel oarecare să depășim 10^{27} biți. Specialistul în neuroștiință Anders Sandberg estimează că potențiala capacitate de stocare a unui atom de hidrogen este de circa patru milioane de biți. Aceste densități nu au fost încă demonstrate, așa că vom folosi estimarea mai conservatoare.

[Notă: Anders Sandberg, „The Physics of the Information Processing Superobjects: Daily Life Among the Jupiter Brains", *Journal of Evolution and Technology 5* (22 decembrie 1999), http://www.transhumanist.com/volume5/Brains2.pdf.]

După cum am discutat mai înainte, 10^{42} calcule pe secundă se pot realiza fără producerea unei călduri semnificative. Folosind la maximum tehnicile reversibile de computare, utilizând formule care generează un nivel redus de erori și permițând o disipare rezonabilă a căldurii, ar trebui să ajungem undeva între 10^{42} și 10^{50} calcule pe secundă.

Domeniul de concepție dintre aceste două limite este complex. Examinarea subiectelor tehnice care intervin pe măsură ce avansăm de la 10^{42} la 10^{50} depășește scopurile acestui capitol. Însă ar trebui să ținem minte că rezolvarea nu va începe pornindu-se de la limita supremă de 10^{50}, de la care să se scadă pe baza diverselor considerente practice. Dimpotrivă, tehnologia își va continua salturile, întotdeauna folosindu-și capacitatea de ultimă oră pentru a progresa spre următorul nivel. Prin urmare, de îndată ce vom ajunge la o civilizație cu 10^{42} cps (la fiecare kilogram), oamenii de știință și inginerii momentului respectiv își vor folosi inteligența în esență vast nonbiologică pentru a-și da seama cum se poate ajunge la 10^{43}, apoi la 10^{44} și așa mai departe. Eu, unul, mă aștept că ne vom apropia foarte mult de limitele supreme.

Chiar și la 10^{42} cps, „computerul portabil suprem" de un kilogram ar putea să realizeze echivalentul gândirii umane din decursul ultimelor zece mii de ani (echivalat la zece miliarde de creiere umane timp de zece mii de ani) în zece microsecunde.

[Notă: Vezi nota de mai sus. 10^{42} cps este o valoare cu 10^{-8} mai mică decât 10^{50} cps, așadar a zece mia parte dintr-o nanosecundă devine 10 microsecunde.]

Dacă examinăm diagrama Creșterii Exponențiale a Computației (Capitolul 2), vedem că acest volum de computație este estimat să fie disponibil contra sumei de o mie de dolari înaintea anului 2080.

CUPRINS

Mulțumiri 7

Introducere 9

Capitolul 1. Experimente mintale despre lume 19

Capitolul 2. Experimentele mintale de gândire 30

Capitolul 3. Un model de neocortex: Teoria minții bazate pe recunoaştere 38

Capitolul 4. Neocortexul biologic 76

Capitolul 5. Creierul vechi 92

Capitolul 6. Abilitățile transcendente 107

Capitolul 7. Neocortexul digital inspirat de biologie 118

Capitolul 8. Mintea ca un computer 172

Capitolul 9. Experimente mintale despre minte 190

Capitolul 10. Legea accelerării veniturilor aplicată creierului 235

Capitolul 11. Obiecții 252

Epilogul 262

Note 267

TOTAL

„Fiind unul dintre cei mai eminenți pionieri ai Inteligenței Artificiale, Ray Kurzweil a creat o nouă carte care ne explică adevărata natură a inteligenței, atât a celei biologice, cât și a celei nonbiologice. Cartea descrie creierul uman ca fiind o mașină capabilă să înțeleagă concepte ierarhice mergând de la forma unui scaun, până la natura umorului. Intuițiile lui deosebit de importante scot în evidență rolul-cheie al învățării, atât pentru creierul uman, cât și în domeniul Inteligenței Artificiale."

Raj Reddy, director fondator al Institutului de Robotică al Universității Carnegie Mellon; laureat al Premiului Turing, decernat de Asociația pentru Mașini de Computație

„Rareori se întâmplă să găsești o carte care să îți pună la dispoziție un conținut unic și atât de încărcat de inspirație în fiecare pagină. *Cum se construiește o minte* realizează însă mult mai mult de-atât. Ray are un asemenea stil de a aborda provocări aparent copleșitoare cu raționamente atât de solide, încât în cele din urmă îl convinge pe cititor că omului îi este perfect posibil să creeze inteligență nonbiologică, o inteligență care să o depășească cu mult pe cea umană. Este o carte vizionară, dar în același timp accesibilă și acaparatoare."

Rafael Reif, președintele MIT

„Dacă te-ai întrebat vreodată cum îți funcționează mintea, citește această carte. Intuițiile lui Kurzweil dezvăluie secrete-cheie care stau la baza gândirii omului și a capacității noastre de a o re-crea. Este o carte elocventă și provocatoare pentru minte."

Dean Kamen, fizician; inventator al primei pompe de insulină la purtător, denumită Aparat de dializă HomeChoice și a sistemului de mobilitate IBOT

Ray Kurzweil este unul dintre cei mai mari inventatori, gânditori și futuriști ai lumii, cu un palmares de treizeci de ani de predicții exacte. Numit „geniul neliniștit" (*The Wall Street Journal*) și „mașina inteligentă supremă" (*Forbes*), Kurzweil a fost ales în topul antreprenorilor de publicația *Inc.*, care l-a descris ca fiind „moștenitorul de drept al lui Thomas Edison".

ISBN 978-973-47-1651-7

PARALELA 45 EDITURA

Preț cu TVA: 29 LEI